

Small-minded government

Last week's debacle in New Orleans highlights failings not just in the Bush administration, but in how the United States chooses to govern itself.

The term 'natural disaster' doesn't really do justice to the scenes that unfolded in the southern United States last week. For a start, the main cause of death in the aftermath of Hurricane Katrina will have been drowning as a result of the flooding in New Orleans that sprang from a widely anticipated failure of the city's flood defences. There is an overwhelming sense that the human calamity that befell the city was avoidable and represents a failure of the US government to protect its most vulnerable citizens.

Much of the blame for the painfully slow reaction to the hurricane has fallen on President George W. Bush, and for good reason. His belated and uninspiring personal response to the crisis has invited widespread criticism. The Department of Homeland Security, the newly created government department that fumbled the early rescue efforts, is viewed as Bush's creation and is ineptly staffed by the president's appointees.

Yet as criticism rains down on the administration, it should be pointed out that several contributory factors that led up to this fiasco preceded Bush's arrival in the White House. These include rampant poverty among African-Americans in New Orleans and other US cities; a systematic failure to build public infrastructure commensurate with America's vast wealth; the habitual creation of dysfunctional government agencies by congressional fiat; and the failure of scientists to successfully convey their concerns to policy-makers.

Previous US flood disasters — notably in Johnstown, Pennsylvania, in 1889 and in the New Orleans area in 1927 — prompted major political upheaval. It is not inconceivable that Katrina will force America's leaders to confront poverty and support public investment in infrastructure. But short of such far-reaching change, the disaster should lead to an immediate re-examination of how the federal government is organized, and how it responds to scientific advice.

The Department of Homeland Security was originally conceived in Congress as a response to the terrorist attacks of 11 September 2001. After initially opposing the idea, Bush co-opted it, removed its most potent aspect (the incorporation of the intelligence agencies) and implemented what was basically an amalgamation of existing

government departments, including the once-admired Federal Emergency Management Agency (FEMA).

According to many observers (see page 174), the reorganization has weakened FEMA and focused its attention on such scenarios as bioterror attacks. The public face presented by FEMA has been diminished, and the agency seems to have retreated from its traditional position at the forefront of disaster response. This weakening has left city and state governments in Mississippi and Louisiana bereft of leadership from the federal government at their moment of greatest need. The lesson is that sweeping reorganizations of government agencies in response to particular crises can have severe adverse consequences.

Knowledge of the risk of a storm-induced flood in New Orleans has been widespread in the scientific community for years, and researchers have sought to improve our understanding of it. Much of this work has taken into account stubborn facts such as the propensity of the poor, the elderly and the sick to ignore evacuation orders.

There seems to be a disconnect, however, between the process that identifies such risks and the people who make the decisions that might manage them. There are indications that many senior politicians — not just President Bush — were simply unaware that the New Orleans flood risk even existed.

River management, meanwhile, has developed into something of a scientific backwater in the United States, some of its practitioners complain. It has also been a subject of bitter political contention — generally between the supporters of the Army Corps of Engineers, which likes to build levees, and environmentalists, who favour marshland conservation and more 'natural' river flow. In the aftermath of Hurricane Katrina, this dialogue-of-the-deaf must end, and the assessment and management of natural risks should be genuinely embraced as a national priority. ■

"Sweeping reorganizations of government agencies in response to particular crises can have severe adverse consequences."

Proteomics' new order

An international organization is finally bringing discipline to the study of cells' sets of proteins.

The international Human Proteome Organisation (HUPO) was launched in 2001 to much scepticism. After all, the proteome — the complete set of proteins expressed by a cell's genome, and modified following expression — is a moving target. Its constituents change continuously according to the conditions to which

the cell is exposed. A 'human proteome project', in contrast to the Human Genome Project, would be a fuzzy, infinite endeavour. What role could there be for an international organization?

Proteomics was in any case enjoying a mixed reputation. The work of a small number of excellent labs was being diluted by vast data dumps of dubious value. The new techniques developed to identify proteins on a large scale were being snapped up by inexperienced users who pumped out data that often proved hard to reproduce. Even in the best hands, the various techniques had very different outputs, making comparison of results between labs difficult. And it was clear to many that the simple cataloguing of long lists of proteins



was not going to carry biology very far forward. The boom was threatened by a bust.

No longer. As last week's fourth annual HUPO meeting in Munich demonstrated, the proteomics community is finally putting its house in order. And HUPO itself is playing an indispensable role. Formally established with a permanent secretariat in Montreal and supported in part by the Canadian government, it has set up committees whose members — the scientific heavyweights of proteomics — select initiatives to come under HUPO's wing. Such projects will now be required to maintain HUPO's standards, and will be supported by HUPO's training and education programmes.

Seven initiatives have already been chosen, including proteome projects in the brain, liver and plasma. Crucially, a major focus will be HUPO's Proteomics Standards Initiative, which is developing standards for data generation and presentation, including standardized formats for databases of the full range of proteomics measurements.

For example, protein-protein interaction data have been generated using both the 'yeast two-hybrid' system and mass spectroscopy, and held in different databases. The initiative has agreed on a common database format that sends users to information in both databases when asked which other proteins a specified protein interacts with.

For mass spectroscopy data, the initiative is developing a list of information that a researcher must provide to accompany a claim that a protein has been identified, including the names of fragments and their individual masses. HUPO may eventually decide that the spectra themselves must be provided.

Two related activities represent important new resources. One is a proteomics database called PRIDE (www.ebi.ac.uk/pride), developed jointly by the European Bioinformatics Institute at Hinxton near Cambridge, UK, and Ghent University in Belgium, which will use the HUPO standardized formats.

The second reflects HUPO's commitment to a broader definition of proteomics in the context of systems biology — using proteomic techniques to follow particular proteins in a biological pathway. Five major protein-interaction databases have agreed to share curating efforts and provide information in a HUPO standardized format, in a consortium called the International Molecular Exchange (IMEX).

Without the umbrella of HUPO, hopes for standardization in proteomics would have been bleak, with researchers being more inclined to use their rivals' toothbrushes than their protocols. HUPO is involving the entire international community in its discussions to ensure consensus, and has already brokered a surprising number of agreements, with journals ready to assist in enforcing standards.

Even as HUPO brings order to chaos, researchers are proving the value that proteomics is bringing to biology, not just in identifying biomarkers useful in medicine, but also in understanding how relatively simple genomic information is transformed into the wonder that is the functioning cell. At the end of the day, proteins, not genes, are the business end of biology. ■

"Without the umbrella of HUPO, researchers would have been more inclined to use their rivals' toothbrushes than their protocols."

Endangered act

Efforts to reform the Endangered Species Act could harm America's most important conservation law.

For three decades, one piece of legislation has been responsible more than any other for protecting land that serves as a habitat for threatened species of plants and animals in the United States. The 1973 Endangered Species Act sought to help ensure the survival of diminishing species by preserving their habitat and aiding the development of science-based 'recovery plans' for their survival. Since its enactment, the act has created and maintained unprecedented natural laboratories for ecologists, and has become a widely admired model for conservation laws worldwide.

As Congress convenes this week, an effort will be made, under the guise of 'reform', to weaken the act. In the final two years of George Bush's presidency, with Republican majorities in the House and Senate, parties that have sought for years to weaken the act — including the mining, oil and logging industries, and property developers — see this as their best chance.

Drafts have been circulating of a proposed revision to the act that would give political appointees in the Department of the Interior far more discretion than they currently enjoy over how the act is exercised. Republican supporters of the bill seek to have it introduced by a Democrat so that the measure will appear bipartisan. But what

is really needed is a show of bipartisan support for the existing law.

Scientific societies and universities need to become involved in this debate now, before the progress that has been made in conservation biology under the act is itself endangered.

The recent apparent rediscovery of the ivory-billed woodpecker in remnants of an old hardwood forest in Arkansas (see page 188) has amply demonstrated the value of the current law. As much as sixty years ago, ecologists were urging the protection of a forest in Louisiana, where what were thought to be the last of the statuesque woodpeckers were known to survive. But the calls were ignored, the forest was chopped, and the ivory-billed woodpeckers disappeared. Such an event would probably have been prevented by the Endangered Species Act. If the ivory-billed woodpecker has indeed re-emerged, it has done so in large part because the act protected the habitat of other threatened species.

"If the ivory-billed woodpecker has indeed re-emerged, it has done so in large part because of the Endangered Species Act."

Everyone loves a rare exotic bird, and the Bush administration has directed some \$10 million specifically towards the habitat and recovery of the ivory-billed woodpecker. But the best way of ensuring the survival of threatened species is through the solid legislative framework that the Endangered Species Act provides, not through single-species projects generated at the whim of politicians. The best outcome for ecology is that the current law be left alone. ■

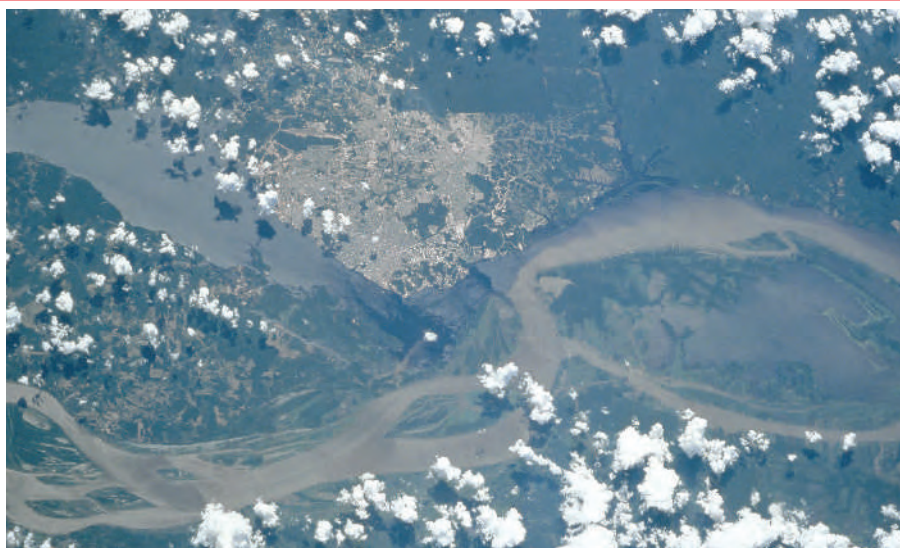
RESEARCH HIGHLIGHTS

Weighing a river

Geophys. Res. Lett. **32**, L16308 (2005)

A GPS station in the Amazon basin has registered flex in the Earth's crust caused by seasonal changes in the weight of the great river.

Michael Bevis of Ohio State University in Columbus and his colleagues compared altitude measurements from a ground station of the Global Positioning System near Manaus in Brazil (pictured) with data from an adjacent river gauge. The daily motions of the ground mirrored that of the water level. The height of the station oscillated by 50–75 mm over an annual cycle, while the water level varied by 10–15 metres. The researchers modelled the data as an elastic response of the Earth's crust and upper mantle to flooding within 200 kilometres of the station.



NASA

MOLECULAR DYNAMICS

Fixed on fractals

Phys. Rev. Lett. **95**, 098106 (2005)

An intriguing result in protein dynamics can be explained by the molecules' fractal nature, find Rony Granek of Ben Gurion University and Joseph Klafter of Tel Aviv University, both in Israel.

Recent experimental studies suggested that the relationship between the relative positions of two points on a particular protein — represented by an autocorrelation function — changes in an unusual fashion. As the protein vibrates, this function decays very slowly, first as a stretched exponential and then as a power law.

Granek and Klafter were able to simulate this behaviour using fractal structures to describe the way in which proteins fill space, showing that the phenomenon is not unique to the protein in which it was observed.

NEUROBIOLOGY

Light sensitive

Neuron **47**, 739–750 (2005)

Adding to the already complex picture of how nerve cells operate, a team reports that action potentials, or spikes, can originate in the tufted dendrites of the retina's output neurons.

The dendrites of retinal ganglion cells receive input currents from other retinal cells. These are thought to be combined in the cell's body, or soma, which controls whether a spike is fired down the cell's axon. But after blocking spike production in the soma, researchers led by Rowland Taylor of Oregon Health and Sciences University in Beaverton, unmasked smaller spikes that

started in the dendrites in response to light. They suggest the dendritic spikes may underlie the ability of some ganglion cells to detect the direction of motion of a light stimulus.

CELL BIOLOGY

Err on a G string

Cell **122**, 633–644 (2005)

Spinocerebellar ataxia type 1 is a neurodegenerative disease, characterized by balance and coordination difficulties, that slowly destroys the Purkinje cells (pictured) in the cerebellum.

It is caused by a faulty gene whose protein product, ataxin-1, contains a longer than normal string of glutamine amino acids. This changes the protein's conformation, and also slows its metabolic breakdown.

Huda Zoghbi from Baylor College of Medicine in Houston, Texas, and her colleagues show that the resulting build-up of abnormal ataxin-1 enhances its interactions with Gfi-1. The transcription factor Gfi-1 is essential for the survival of Purkinje cells in adults, and the researchers find that ataxin-1 promotes its destruction.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

EPIGENETICS

Mapping yeast

Cell **122**, 517–527 (2005)

Improved high-resolution techniques have produced the first genome-wide map of gene regulation in yeast.

Richard Young and his colleagues at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, applied techniques including DNA microarrays containing 40,000 DNA probes to the yeast genome. The resulting map details the location of the histone proteins that organize the structure of DNA. It also shows where the histones have been modified by the addition of chemical groups (that is, by acetylation or methylation), which in turn switches genes on and off.

MEDICINE

Two for one

J. Exp. Med. **202**, 1–10 (2005)

A new vaccine can protect mice against two different fungal pathogens. A team headed by Antonio Cassone of the Italian National Institute of Health in Rome designed the vaccine to sensitize the animals' immune system to a type of large sugar molecule found in many fungal cell walls — β -glucan.

Conjugating the sugar with a carrier protein culled from a diphtheria toxin increased the vaccine's potency. More than 60% of vaccinated mice survived for 30 days following exposure to the fungal pathogen *Candida albicans*, compared with 20% of a control group. And the vaccine nearly tripled the survival rate of mice dosed with *Aspergillus fumigatus* spores.

MOLECULAR BIOLOGY

Immortal coils

J. Cell Biol. **170**, 721–732 (2005)

Since the 1970s, scientists have reasoned that it makes sense for stem cells to keep the original DNA when they divide because it contains the fewest replication errors. Philip Karpowicz of Canada's University of Toronto and his colleagues now offer evidence from mouse neural stem cells to support this debated theory, known as the immortal strand hypothesis.

The team labelled the DNA in a neural stem cell with a compound called BrdU. This cell then divided to form a cluster of cells. The BrdU-labelled DNA appeared in fewer cells than would be expected if the strands from the original chromosomes had dispersed randomly. However, such asymmetric segregation was not seen in other types of stem cells from mice, suggesting the neural stem cells represent a special case.

CANCER

Multitasking

Genes Dev.

doi:10.1101/gad.1339905 (2005)

The tumour-suppressor gene p53 can be transcribed in six different ways, report researchers led by Jean-Christophe Bourdon and David Lane of the University of Dundee, UK.

The researchers showed that expression of the resulting protein products varies between tissues, and speculate that the interplay between these different protein isoforms is key to p53's normal role, and could even underpin the progression of cancers.

The team found that, like the related genes p63 and p73, the human p53 gene contains an internal promoter region. This means transcription can start in the middle of the gene, with partial transcription giving rise to the short isoforms of the p53 protein. The same internal promoter is conserved across species, indicating its importance. But what regulates the promoter is unclear.

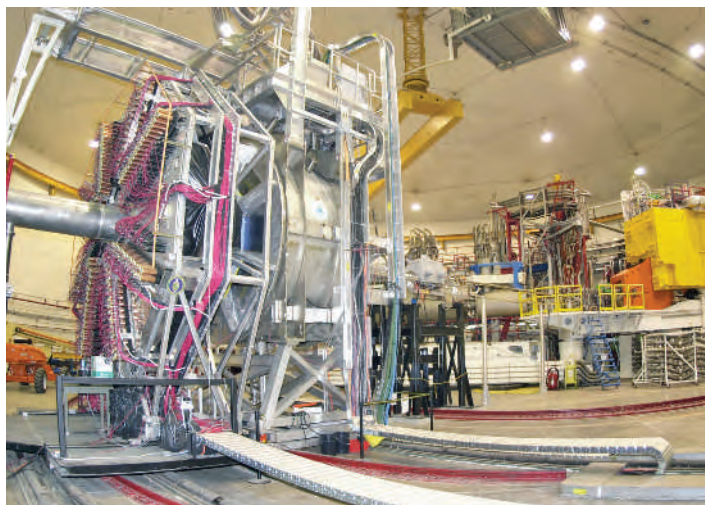
MEDICINE

Blood relations

J. Clin. Invest. doi:10.1172/JCI24177 (2005)

The myelodysplastic/myeloproliferative diseases are a group of disorders in which

blood stem cells fail to develop properly — leading to abnormal development of blood cells from bone-marrow precursors. Studying these diseases has been extremely difficult because there has been no single gene associated with them, or animal model of the disorders. Now researchers led by Carlos Martinez-A of the National Center for Biotechnology in Madrid and Miguel Campanero of the Autonomous University of Madrid have tackled both these problems at once. They identify a gene called *Dido* whose altered expression causes myelodysplastic/myeloproliferative-like diseases in mice, and show that altered expression of the gene also occurs in people with the diseases.



JEFFERSON LAB

PARTICLE PHYSICS

Ups and downs

Phys. Rev. Lett. **95**, 092001 (2005)

Inside the proton is a flurry of activity: three permanent quarks, two ups and a down, are surrounded by quark–antiquark pairs that zip in and out of existence. These fleeting ‘sea quarks’ — primarily strange quarks — contribute to the proton's magnetism and charge.

Now the G0 Collaboration, which is based at the Jefferson Lab in Virginia, Newport News, present a detailed picture of this strange sea. This was created by scattering fast, polarized electrons off protons within a hydrogen target (G0 detector pictured). The data suggest that strange quarks contribute about 5% of the proton's magnetic moment. This is consistent with previous experiments, and highlights a discrepancy with the latest theoretical prediction, which is ten times smaller and has the opposite sign.

JOURNAL CLUB

Jennifer Doudna

University of California,
Berkeley

A biochemist exchanges her ideas about molecular motion.

Proteins that convert chemical energy into physical motion fascinate me because they turn the mundane into the extraordinary. They use the simple act of adding water to nucleotides, such as GTP, to drive everything from molecular shape changes to the movement of molecules within cells.

One of their most familiar roles is in helping the ribosome to stitch together amino acids into protein chains. I learnt about this process as an undergraduate, and followed the field through the explosion of interest sparked by high-resolution structures of the ribosomal subunits — so I assumed that I knew the basics.

In the classical model of protein synthesis, the ribosome ratchets along messenger RNA, gathering the amino acids encoded in the RNA sequence. The growing protein chain is extended when the elongation factor EF-G binds to the ribosome, bringing with it a molecule of GTP. Hydrolysis of the GTP drives the addition of the next amino acid by propelling the ribosome along the messenger RNA, which in turn triggers the used EF-G to dissociate. Textbook simple, right?

Perhaps not. A recent paper in the *Journal of Biology* (A. V. Zavialov, V. V. Hauryliuk and M. Ehrenberg **4**, 9; 2005) has made me reconsider this model. It provides evidence that EF-G binds to the ribosome attached to a nucleotide known as GDP, rather than the energy-providing GTP. The ribosome then catalyses the exchange of GDP for GTP, which suggests that it plays a previously unrecognized part in accelerating protein synthesis.

The findings focus new attention on the role of GDP and GTP exchange in driving molecular motion — and show that an established field can still spring surprises.

SPECIAL REPORT

After the flood

Academic experts say they were all too aware of the devastation that would claim New Orleans and its surroundings in the wake of a fierce hurricane. Could they have done any more to convince politicians of the need to protect the city?

Nothing about last week's hurricane and the subsequent flooding of New Orleans should have come as a surprise. Experts knew such a storm would come at some point. They knew the coast's natural defences were degraded; they knew the levees were not designed for anything stronger than a category-3 storm; and they knew that a significant proportion of the population — the poorest and weakest — would not evacuate.

The science was all there, but apparently the planning was not. As the United States reels from one of the worst disasters in its history, scientists are trying to work out why policy-makers were unable to cope when experts knew so much about what was bound to happen.

Public officials criticized for mishandling the emergency have claimed that Katrina was simply too powerful a force to resist. New Orleans hadn't been hit by such a large storm since Hurricane Betsy in 1965 — at category 5, Katrina was still on the highest possible level of the Saffir-Simpson scale mere hours before landfall on the morning of 29 August (see 'The power of Katrina'). Breaches in the city's levee system then turned a bad situation into a catastrophe, flooding an area of more than 400 square kilometres with water from Lake Pontchartrain and trapping tens of thousands of people in a swiftly escalating crisis (see 'Threat of disease').

But the dire consequences of a large hurricane striking New Orleans have been predicted



Hurricane Katrina pounded the Gulf coast with winds of more than 200 kilometres per hour.

for years, precisely because of the risk of flooding (see *Nature* 431, 388; 2004).

It is public knowledge that the sandy barrier islands and marshy bayous that used to protect the Louisiana coast from storms and hurricanes are eroding as dams and levees hold on to the silt that usually rebuilds them. The marshes are disappearing at a rate of more than 60 square kilometres a year. In 1998, a document called Coast 2050 was drawn up by state officials calling for restoration of the wetlands. However, the full cost of the project is estimated as \$14 billion, and the state has made little progress in persuading

federal government to give it more than a tiny fraction of the request.

It has also long been known that the system of raised levees and floodwalls that keep New Orleans dry are only designed to withstand hurricanes up to category 3. A project looking at upgrading the system is in the works, but after five years it is still in the pre-study phase. To be ready for Katrina, "we would have had to start working on category-5 twenty years ago", says Alfred Naomi, a senior project manager for the Army Corps of Engineers, which maintains the levee system.

A hurricane strike at some point was inevitable, say researchers. And they argue that there was also no excuse for not realizing the potential scale of the disaster.

Rick Luettich of the University of North Carolina in Chapel Hill helped to develop the Advanced Circulation Model that is used by

Louisiana State University (LSU) in Baton Rouge and the Army Corps of Engineers. The model has become increasingly accurate at predicting storm surge — a wind-driven rise in sea level — even at small scales. Luettich says

it got the effects of Katrina "about right". Days ahead of the storm's arrival, computer simulations of the expected surge showed that water would probably overflow levees, flooding the city, which lies below sea level.

"It's a valuable tool," Luettich says. "Where we've had less success is in getting people to take it seriously and modify their behaviour based on it."

"Academia tends to be discounted," agrees Ivor van Heerden, director of LSU's Center for the Study of the Public Health Impacts of Hurricanes. "But we called this 100% right."

Politicians were slow to act on warnings from the models and related casualty simulations (see 'Counting the dead'). New Orleans' mayor, Ray Nagin, did not issue a mandatory evacuation order until the day before the hurricane hit — too late for many. An estimated 80% of the city's 470,000 residents evacuated using extra highway lanes that had been opened for the emergency — one of the few parts of the disaster plan that worked well. But that still left roughly 100,000 people in the city.

"Academia tends to be discounted. But we called this 100% right."

THE POWER OF KATRINA

Katrina was the third most intense storm ever to make landfall in the United States, with a central pressure of 918 millibars. The storm affected an area of about the size of Britain, and the maximum storm surge was 10 metres, recorded in Biloxi, Mississippi.

The intensity of the storm is due partly to this year's high sea surface temperatures — in early August the Gulf of Mexico was 2–3 °C warmer than usual for the time of year. That led to perfect conditions for the formation of Katrina, and provided the energy that caused such destruction when the storm hit land. Katrina sucked so much heat from the gulf that water temperatures dropped dramatically after it had passed,

in some regions from 30 °C to 26 °C.

Forecasts don't augur well for the rest of the season. The hurricane risk usually peaks in August and September, and this year has already been exceptional for the number and intensity of storms, according to London-based firm Risk Management Solutions, which models disaster risks for the insurance industry. Sea surface temperatures in the central Atlantic are likely to remain at least 1 °C above normal until October, according to forecasts by the US National Oceanic and Atmospheric Administration and Colorado State University, so the risk of hurricanes is likely to remain high until then.

Quirin Schiermeier



THE AFTERMATH OF HURRICANE KATRINA
Read ongoing coverage of the New Orleans floods and their implications.
www.nature.com/news

ORBIMAGE

V. LAFORET/AP

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

This came as no surprise to LSU sociologists Jeanne Hurlbert and John Beggs. In 2004, their analysis of survey data suggested that 21% of the population would stay in their homes during a hurricane, and that 32% would remain in the area. People in poor health were especially likely to stay, as were those suffering from depression, disabilities and other life stresses — Hurlbert characterizes them as “people who are already not coping”.

Based on these data, Devyani Kar of LSU's Coastal Studies Institute was in the process of mapping New Orleans with a Geographic Information System database that includes variables such as income level and access to transportation, to determine which neighbourhoods were likely to have the most flood victims. Her study was not yet finished by the time Katrina struck.

Even so, there should have been strategies in place to ensure that the most vulnerable populations were evacuated, says Ilan Kelman, an expert on disasters at the National Center for Atmospheric Research in Boulder, Colorado. Public officials should be “entirely proactive” he says, and not rely on news reports to convey

THREAT OF DISEASE

Most of the deaths from Hurricane Katrina will have been caused by drowning — but the danger to those in the area was soon followed by a lack of food, water, shelter and sanitation.

“The immediate need is to provide shelter for all the displaced people,” says Eric Weiss, professor of emergency medicine at Stanford University Medical Center in California. “New Orleans is like the biggest wilderness in the United States right now. We need to get people to adequate shelter so they're not baking in the sun, which predisposes them to dehydration. The second thing they need is drinking water. There's a saying in wilderness medicine that people can live three hours without shelter, three days without water and three weeks without food.”

Experts agree that if people are evacuated quickly, the threat of disease is low. “The most important thing is to get people out,” says Louisiana state epidemiologist Raoult Ratard, speaking from a temporary operations centre in Baton Rouge. “In developed countries, if you get the population into shelters with fairly safe

environmental conditions, they are going to be OK.”

But by the end of last week, conditions were looking less and less like those of a developed country. Up to 20,000 people were trapped in hellish surroundings at the Superdome in New Orleans for a full five days after the flood waters rose, before being evacuated to other shelters. Speaking on 2 September, Ratard said that there were no reports of major disease outbreaks, but added that no surveillance officials had yet entered the Superdome or the flooded city.

Dead bodies are not thought to pose a large disease risk. Typhoid and cholera are also unlikely to be major threats, because the organisms that cause them are not endemic in New Orleans water. But there will be a major risk from diarrhoeal diseases, respiratory-tract infections and mosquito-borne illnesses, especially West Nile virus. The hurricane has temporarily cleared away the birds and mosquitoes that carry the virus, but Ratard warns that the virus could surge back in the weeks to come.

Erika Check

COUNTING THE DEAD

As rescue operations continue in New Orleans, a grim question looms: how many lives will Hurricane Katrina ultimately claim?

When federal and state agencies conducted an exercise called Hurricane Pam in July 2004, IEM, the Baton Rouge contractor that developed the hurricane disaster plan for New Orleans, estimated that 60,000 people would die in southeastern Louisiana alone. The initial estimate was 80,000, but Louisiana State University researchers say this figure met with such scepticism from government emergency planners that it was reduced.

The similarities between Katrina and the Pam simulation are “eerie”, says IEM president Madhu Beriwal. In both cases, water flows over

or through the protective levees, flooding parts of the city with up to 7 metres of water. In both cases, drowning is the leading cause of death, depending greatly on the depth of the water.

Ezra Boyd, who models such emergencies at Louisiana State University, used data from past hurricanes to estimate the number of deaths that would be caused by a category-3 storm with a 5-metre storm surge. He came up with a similar number: 72,000 deaths in the greater New Orleans area alone.

Neither estimate includes the more than 350,000 people who live in the coastal Mississippi counties. These areas were not inundated like New Orleans, but experienced higher winds than the city and greater storm

surges — more than 8 metres in places. If 30% of the population stayed behind, more than 100,000 people may have suffered Katrina's fury.

A general rule of thumb, according to Han Vrijling of Delft University of Technology in the Netherlands, is that fatalities from flooding can amount to 1% of the affected population. Even that conservative estimate could yield as many as 5,000 deaths from Katrina.

As *Nature* went to press, workers had no death-toll estimates. But the researchers caution that casualties are highly dependent on the individual circumstances. Besides, says Beriwal, “it hasn't finished happening. We are still trying to save lives.” **Tony Reichhardt**

information. Warnings should be tailored to people of diverse backgrounds. “It means setting up a website, and it means talking one-on-one with a homeless person on the street,” he says. And for those who remain to ride out a hurricane, emergency supplies should be pre-positioned, to speed up rescue operations after the storm passes.

State of unreadiness

David Ozonoff, an environmental epidemiologist at Boston University in Massachusetts, agrees that there is no excuse for New Orleans' refuges being so poorly set up for survivors. “This should have been anticipated,” he says. “People moving is not an unknown and unsolved problem. It can be very difficult, but if you've thought it out ahead of time you should be prepared.”

Ozonoff believes part of the problem is that after the terrorist attacks of 11 September 2001 a shift in federal priorities crippled the US public-health and disaster-response agencies. The US Federal Emergency Management Agency was taken under the umbrella of the Department of Homeland Security, which is supposed to coordinate local and national responses to all kinds of disasters. The hope was that efforts to prepare for terrorist attacks would improve preparations for natural disasters; Ozonoff argues that the response to Katrina proves the approach wrong. “If there was ever an episode — other than a bioterrorist attack — that should have demonstrated that money put into public-health preparedness

is effective, then this was it,” he says.

He criticizes the Department of Homeland Security for focusing on expensive gear to deal with unlikely bioterrorist attacks instead of on the less dramatic but crucial task of coordinating responses to more realistic scenarios. “The whole bioterrorism prevention effort has brought nothing in the way of preparation, and has left public-health departments in much worse shape than before they got the money,” he asserts.

Others say scientists should take part of the responsibility. “Folks have talked about this scenario for decades, yet I've watched George Bush senior and Bill Clinton both comment that no one could have anticipated this sort of event,” says Roger Pielke, director of the Center for Science and Technology Policy Research at the University of Colorado, Boulder. “That raises some real questions for the academic and scholarly community. What does it mean if scholars are aware of something with practical importance, but it doesn't get to the people who can take action?”

“I knew from the models how bad it was going to be.”

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Studies predicted that 32% of residents would fail to leave New Orleans.

Pielke argues that scientists need to move away from a ‘loading dock’ approach where they simply put out raw information for anyone who wants to use it. Instead they should tailor their research to practical needs, he says. “There's a real challenge of making knowledge useful. It is not something that the academic community is engaged in as a matter of policy.”

His plea is echoed by van Heerden: “Academia gives more credit for journal publications than for helping a hospital prepare for a crisis.”

Getting hands on

“Universities aren't particularly well organized to support applied research,” adds Pielke, “but events like Katrina and research that languishes in journals should motivate policy-makers to demand more from the scientific community.” Pielke concedes that attitudes within the scientific community are beginning to change, with agencies such as the National Science Foundation supporting some practical work: “There's a trend in the right direction.”

But for researchers already carrying out such applied projects, the mood is one of frustration. LSU's Nedra Korevec, for example, has studied the scenario of a category-4 storm striking New Orleans. She managed to convince many of her family and friends to move in with her in Baton Rouge before the hurricane hit, and is now housing ten people. “I knew from the models we ran and the work I'd done how bad it was going to be,” she says. “We do the research and we try to make things happen, but then we have to hand the ball to delegations and lobbyists.” ■

Tony Reichhardt, Erika Check and Emma Maris

Health centres and labs left reeling by Katrina

In the aftermath of last week's hurricane, scientists from research facilities and medical centres in the area are counting the cost — to patients, buildings and their research.

Hardest hit are three medical facilities — Charity Hospital, and the health science centres at Tulane University and Louisiana State University (LSU), all of which are located close together in downtown New Orleans.

At the Tulane centre, Paul Whelton, physician and senior vice-president for health sciences, says that for four desperate days after the waters rose, patients were kept alive with hand-pumped ventilators. Surgeons swam through dirty flood waters to reach nearby hospitals, he adds, while faculty scientists kept freezers functioning and some 5,000 research rodents alive on the slimmest of supplies.

"It was truly heroic," says Whelton, who was evacuated along with the last of the staff and patients on 2 September. "I couldn't be more proud of the faculty, staff and administration."

The nearly 13,000 undergraduate students at Tulane evacuated their campus before the hurricane hit on 29 August, but the science staff stayed behind to look after patients and research facilities. Despite the loss of power when the auxiliary generators ran out of fuel, no staff or patients were lost, says Whelton, and about 250 patients were flown out by helicopters provided by the Hospital Corporation of America, which manages the facility, before military helicopters arrived.

At nearby Charity Hospital, which cares for New Orleans' poorest, the situation was grimmer. "It was bad," says Whelton. "They had dead laying in the stairwells." About 30 critically ill patients were moved in small boats to Tulane to be flown out. "Faculty and nurses rowed them over. Some were on portable respirators. Staff worked in shifts hand-pumping air to them," Whelton says.

Whelton had no information about LSU's Health Science Center, situated a few blocks away, although Norka Ruiz Bravo, the National Institutes of Health's deputy director for extramural research, says he finally heard from officials there on 4 September. Attempts are being made to move some of the centre's core research specimens to facilities in Baton Rouge or Pennington.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A. J. SISCO/UP/NEWS.COM

Patients evacuated by helicopter get treatment at Louisiana State University in Baton Rouge.

Before the last staff left the Tulane facility, they topped up nitrogen tanks to keep their cell lines, bacterial cultures and other specimens frozen. But some research animals are likely to have been lost. As *Nature* went to press, Tulane personnel were planning to return by truck through the 1–2 metres of water that still surrounds the hospital, to do a damage assessment and try to rescue irreplaceable cell lines, such as those from the stem-cell and gene-therapy facilities. "We are hoping for the best, but fearing for the worst," says Whelton.

Facilities elsewhere escaped more lightly. The Stennis Space Center in Mississippi lost the roof from its administration

building, and other buildings were damaged by water. "The good news is there were no staff fatalities or injuries," says NASA spokesman David Steitz. "But it looks like most of the workforce lost their homes."

NASA's Michoud Assembly Facility, in the coastal lowlands east of New Orleans, could only be reached by helicopter or boat for several days after the storm. The space shuttle's external fuel tanks are made there, but preliminary indications are that the tanks were not damaged, although damage to the buildings could still delay future flights.

Meanwhile, the Laser Interferometer Gravitational Wave Observatory at Livingston, and the Tulane National Primate Research Center, north of New Orleans, did not suffer significant damage. Conferences due to be held in New Orleans' convention centre have been cancelled, at least up to December.

Rex Dalton

"It looks like most of the workforce lost their homes."



DARK MATTER HIGHLIGHTS EXTRA DIMENSIONS

Three new 'directions' may explain astronomical puzzle.
www.nature.com/news

NASA

Drug agency accused of political bias

WASHINGTON DC

A top official at the US Food and Drug Administration (FDA) resigned last week in protest at the agency's failure to make a 'morning after' contraceptive more readily available. Her departure fuelled charges that the agency has surrendered its scientific independence to conservative pressure from the administration of President George W. Bush.

Susan Wood, who directed the FDA's Office of Women's Health, wrote in a 31 August e-mail to her colleagues: "I can no longer serve as staff when scientific and clinical evidence, fully evaluated and recommended for approval by the professional staff here, has been overruled."

Wood told *Nature* that the FDA's recent decision to delay making the 'morning after' pill Plan B (levonorgestrel) available over the counter bore the hallmarks of political interference. "It was very unusual for the commissioner of the FDA to essentially overrule the recommendation of all of the professional staff," she says.

The drug's manufacturer, Barr Pharmaceuticals in Woodcliff Lake, New Jersey, originally applied in 2003 to get Plan B's prescription-only status changed so that it would be available over the counter. The FDA rejected this in 2004, going against its own staff and an outside panel of expert advisers, who voted 23–4 in favour of

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Campaigners have found FDA delays over the 'morning after' pill Plan B hard to swallow.

the change. Barr then followed the FDA's advice by resubmitting the application, this time asking for the contraceptive to be available over the counter only to women over 16.

A promised decision on that application was delayed late last month, when FDA commissioner Lester Crawford announced that Plan B presented "unique" regulatory obstacles.

Activists who believe that life begins at conception have vehemently opposed making access to the contraceptive easier. "Thank

goodness there is now one less political activist at the FDA who puts radical feminist ideology above women's health," says Wendy Wright, senior policy director at Concerned Women for America, a conservative activist group in Washington DC.

Pressure has also come from the other side. Two Democrat senators this spring blocked Crawford's nomination as FDA commissioner until they were guaranteed a final decision on Plan B's over-the-counter availability by 1 September (see *Nature* 436, 442; 2005). The two senators, Hillary Clinton (New York) and Patty Murray (Washington), expressed outrage over the new delay, calling it "a breach of faith".

"The president now has his FDA administrator, but the American public still does not have an answer on Plan B," they wrote in a letter on 30 August to Michael Leavitt, the Secretary of Health and Human Services.

In a similarly angry editorial, published online last week by *The New England Journal of Medicine* (doi:10.1056/NEJMp058222), editor-in-chief Jeffrey Drazen and others wrote that the FDA's actions have "made a mockery of the process of evaluating scientific evidence, disillusioned many of the participating scientists both inside and outside the agency [and] squandered the public trust." ■

Meredith Wadman

WOMEN'S CAPITAL CORP./GETTY

Poles lose out as ozone levels begin to recover

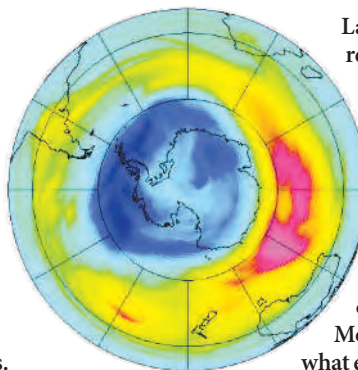
There are mixed tidings for Earth's protective ozone layer this year. Although the seasonal Antarctic ozone hole has already swollen to the size of Europe, the ozone layer worldwide seems to be on the road to recovery.

Ozone molecules in the atmosphere shield Earth from harmful ultraviolet radiation, but man-made chemicals called chlorofluorocarbons (CFCs) eat away at the protective layer. Last week, the European Space Agency's Envisat satellite recorded an exceptionally large seasonal ozone hole forming over the South Pole (shown right in dark blue). Only the holes of 1996 and 2000 have been larger at this time of year.

But everywhere except the polar regions, the ozone layer seems to have stopped thinning as early as 1996, a study now suggests. CFC use was phased out by the 1987 Montreal Protocol, and researchers have been monitoring ozone levels for signs of recovery.

The latest study relies on a sophisticated regression analysis of ozone measurements from satellite- and ground-based instruments taken between 1978 and 2002 (G. C. Reinsel *et al.* *J. Geophys. Res.* 110, D16306; 2005). It found evidence for a modest but lasting ozone recovery, particularly at high latitudes.

"Our findings do offer hope," says Elizabeth Weatherhead, an environmental researcher at the University of Colorado at Boulder and one of the study's co-authors. "But we are still in a depleted state, and there has also been very little improvement in lessening of ultraviolet radiation since the mid-1990s."



Last spring, scientists reported the biggest ozone losses ever recorded over the Arctic, probably due to unusually low winter temperatures (see *Nature* 435, 6; 2005). Experts are divided as to whether the turnaround in ozone levels can be attributed to the Montreal Protocol, and if so to what extent. But most researchers

agree that because CFCs last for a long time in the atmosphere, ozone levels should have continued to fall until the end of the last century. Some scientists therefore argue that if ozone destruction peaked as early as 1996, other changes in atmospheric dynamics must also be at work. ■

Quirin Schiermeier

KNMI/ESA

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Sea change: Fijian fishing families are among beneficiaries of recent self-help aid projects that protect their ecosystems.

Ecology is key to effective aid, UN told

Support is growing for the idea of linking aid with environmental protection. A new report proposes that enabling local communities to manage natural resources is the key to tackling poverty. In response, development experts are calling for a more scientific approach to deciding how aid money should be spent.

The Wealth of the Poor, backed by the United Nations (UN) and the World Bank, says that governments have ignored the 750 million rural people who make up three-quarters of the poor, and that development in areas such as China and India has largely bypassed rural areas. Marine and agricultural resources around the world are being depleted as small-scale farmers struggle to feed their families.

"Development strategies represent the priorities of capital cities and finance ministers," says Jonathan Lash, president of the World Resources Institute in Washington DC, the environmental think-tank that wrote the report.

It follows a global environmental survey that concluded, in March, that ecosystems around the world are in decline and will continue to be so unless their value is factored into financial thinking (see *Nature* **434**, 547; 2005).

One solution, suggests the report, is for development agencies to focus on wealth created when people work together to manage local ecosystems. During the past decade, for example, communities in Fiji have reversed a decline in marine catches by confining fishing

to restricted areas. In northern Tanzania, land reforms helped villagers band together to succeed where overseas agencies had failed, reforesting around 3,500 square kilometres of badly degraded land that now provides fuel and food.

But the report includes few such successes. "If you tried to find 50 more you couldn't," says Lash. Organizations such as the World Bank do not consider the 'ecosystem services' that underpin the successful case studies, he says.

"That's a correct and important point," says Jeffrey Sachs, director of the Earth Institute at Columbia University in New York. If ecosystem services are

not taken into account, he says, aid agencies go on supporting such schemes as fisheries that raise incomes in the short term but reduce community resources as fish stocks shrink.

Sachs adds that development agencies fail to involve researchers in schemes that focus on ecosystem services. "There is lots of good science available, but very little is tapped for public policy," he says. "We have two cultures. Most people who have trained in economics are not trained in science."

Scientific input, say the few researchers who have taken part, can help people see if changes are working and test out future options.

In KwaZulu-Natal, South Africa, for example, scientists helped local women design

experiments to assess the ecological impact of different clam-harvesting strategies.

"The point is for scientists to help communities gather the information they need to manage their marine resources better, rather than telling them what to do," says Bill Aalbersberg, an applied ecologist at the University of the South Pacific in Fiji, who helped communities monitor the impact of no-catch zones.

Some development agencies say they are taking the message on board. The UK Department for International Development, for example, last year appointed a chief science adviser and announced plans

to increase its research budget by more than £50 million (US\$90 million) to £135 million in 2007. It says it is working with environmental organizations such as the WWF to ensure that the work it supports involves local stewardship of natural resources.

Lash hopes that the report will influence the UN's Millennium Development Goals, a set of targets agreed in 2000. The goals, which have been criticized for treating poverty and the environment as separate issues, are to be discussed in New York next week. But Lash fears his plans may have been derailed by US ambassador John Bolton, who has requested revisions that could water the goals down. ■

Jim Giles



BATTLE IN THE BRAIN
PREDICTS RISKY ACTIONS
 Pleasure and anxiety centres
 decide when a safe bet
 beats a dicey one.
www.nature.com/news

Budget plans hint at lean times ahead for Japanese research

TOKYO

Next year promises to be another tight one for Japan's finances, and for the first time in years, basic research may feel the pinch too.

Last week saw government departments submit their initial budget requests for 2006. As for 2005, the Ministry of Finance is seeking to cut 1.5% from the overall budget — but this time it has suggested that research funding will have to play its part in the cutbacks.

Japan needs to reduce overall government spending because of its huge national debt, which at ¥900 trillion (US\$8.2 trillion) amounts to 170% of its gross domestic product. But the government had so far exempted the research budget from cost-cutting, in an effort to boost global competitiveness.

Research, including the main grant system, accounts for a third of Japan's ¥3.6 trillion science and technology budget. Although this budget has been declining gently for the past five years, money for research projects has been growing steadily since 1989. But this year "the overall budget has become tighter," says Tsuyoshi Maruyama, head of the Science and Technology Policy Bureau at the education ministry, and the finance ministry has suggested a cut of 2–3% to the research budget.

To cope with such constraints, the government wants to prioritize how money allocated for research is spent. It plans to focus on large-scale projects such as space missions and supercomputers, which have lost out in recent years to areas such as nanotechnology and life sciences.

Among the big-ticket projects on the educa-

tion ministry's shopping list for funds is the Global Earth Observation System of Systems. This ten-year international project will create a unified Earth observation system (see *Nature* **433**, 789; 2005), and the ministry is seeking ¥27.6 billion for Japan's contribution to it next year. Other areas that would benefit under the budget request include the development of advanced proteomics analysis technologies and RNA research.

But many scientists are worried that this new focus on big projects will absorb the lion's share of the money, meaning cuts for other areas. One senior official at the education ministry says he fears that cash for life sciences overall could be cut by about ¥13.5 billion. All of that would come from research projects, he adds, as it is difficult to reduce personnel costs.

Yoshiki Hotta, a biologist and head of the Research Organization of Information and Systems, says he understands the need to streamline costs. But he adds that he is disappointed that the tighter budget meant the education ministry didn't pass on many of his organization's research proposals to the finance ministry. "Even though we want to present new ideas, we can't," Hotta says. "That will discourage researchers and dampen Japan's science."

The final decision on the research budget won't be taken until December, and the finance ministry may yet prove to be flexible. It could, for example, allocate science a slice of a ¥100-billion special budget that the government will set aside to finance important policies.

Ichiko Fuyuno



Scaling down:
 researchers in Japan
 may have to tighten
 their belts next year.

Chernobyl: poverty and stress pose 'bigger threat' than radiation

The Chernobyl nuclear accident in 1986 will lead to far fewer deaths than originally thought, according to a report from the United Nations.

Some 4,000 people — including emergency workers and residents of the most contaminated areas — could eventually die of factors linked to radiation exposure, the report says. Earlier estimates had ranged widely but regularly suggested that there could be many tens of thousands of deaths (see *Nature* **351**, 4; 1991).

"The effects on public health were not nearly as substantial as had at first been feared," says Michael Repacholi, head of the radiation programme at the World Health Organization in Geneva.

Repacholi was among more than 100 scientists, economists and health experts who worked on the 600-page document, which aimed to summarize the available scientific data on the accident and the countries most affected by it: the Ukraine, Belarus and Russia. The report was due to be released by the Chernobyl Forum at a meeting in Vienna this week.

On 26 April 1986, one of four reactors at the Chernobyl nuclear power plant in the Ukraine suffered a series of explosions and a meltdown — the worst nuclear accident in history. Radioactive fallout contaminated more than 200,000 square kilometres of Europe, leading to the eventual relocation of more than 350,000 people.

So far, the report says, fewer than 50 deaths have been directly attributed to radiation exposure — most of them rescue workers who died of acute radiation syndrome shortly after the disaster. The authors estimate that the incidence of radiation-induced cancer rose by only about 3% in the affected areas.

They add that the thousands of children who contracted thyroid cancer after the accident are likely to have a 99% survival rate, higher than the 80–85% previously thought.

Poverty and mental-health problems, such as stress, depression and anxiety, pose a much greater threat to the local communities than radiation, the report concludes. It argues that future aid should focus on improving the healthcare system and promoting local economic development.

Valeska Stephan

Mutating stem cells provide culture shock

Embryonic stem cells that are cultured in the lab have been found to accumulate an alarming array of genetic changes. The discovery, published this week in *Nature Genetics*, suggests that such cells can undergo only a limited number of divisions before they become unsafe for therapeutic purposes. It will also increase pressure on the US government to relax the rules that prevent researchers from using federal funds to generate new lines of human embryonic stem cells.

In June, at the meeting of the International Society for Stem Cell Research in San Francisco, researchers revealed that cultured embryonic stem cells accumulate 'epigenetic' modifications — changes, such as the methylation of their DNA, that can alter their behaviour (see *Nature* 436, 9; 2005).

The latest study confirms this finding, and shows that cells cultured over long periods also develop genomic changes that might be associated with cancer (A. Maitra *et al. Nature Genet.* doi:10.1038/ng1631; 2005). "The mutations we are finding are a much bigger problem because they are a permanent change," says Aravinda

Chakravarti of the Johns Hopkins University School of Medicine in Baltimore, Maryland, who led the research.

African telescope opens window on southern sky

The Southern African Large Telescope (SALT), the biggest optical telescope in the Southern Hemisphere, has taken its first images.

Located near the Kalahari Desert, the 11-metre telescope was built over five years by an international consortium of more than a dozen institutions. Despite its size,



The 11-metre Southern African Large Telescope has focused on the Lagoon nebula.

SALT cost only US\$30 million, thanks in part to having a primary mirror made up of 91 fixed hexagonal segments. The mirror cannot move to track celestial objects across the sky, but can only study what comes into its field of view.

Astronomers hope that SALT, which uses a camera built in South Africa and digital spectrographs from the United States and New Zealand, will provide a unique window on the southern sky.

Salted roads create hazard for freshwater animals

Salting roads to make them passable in icy weather is making freshwater streams dangerously salty, according to new research from the United States.

An interdisciplinary team has studied watersheds in Maryland, New York and New Hampshire over the past three decades. During that time, rivers and streams have become substantially saltier; in some cases, in winter, the water contained chloride concentrations nearly one-quarter that of sea water.

If salinity continues to increase at this pace, much of the water would become undrinkable and poisonous to aquatic

creatures, the scientists report (S. S. Kaushal *et al. Proc. Natl Acad. Sci. USA* doi:10.1073/pnas.0506414102; 2005).

Humans proposed as the source of mad cow disease

The first case of mad cow disease may have come from human remains in Bangladesh that were ground up into cattle feed, researchers suggest.

Many experts in the field view the idea with scepticism, saying the evidence remains circumstantial. But there are few alternative explanations for the origin of bovine spongiform encephalopathy (BSE), widely known as mad cow disease, so the theory has attracted attention.

Alan Colchester of the University of Kent in Canterbury, UK, and his daughter Nancy Colchester, of the University of Edinburgh, point out that Britain imported hundreds of thousands of tonnes of whole and crushed bones and animal carcasses during the 1960s and 1970s. These were used for fertilizer and to feed livestock.

Nearly half of the imports came from Bangladesh, where peasants gathering animal materials may have also picked up human remains. These could have been

Stored artefacts get the chance to shine

A treasure trove of scientific artefacts, including the safety lamp of railway pioneer George Stephenson (shown here), is to be freed from basement stores to go on display at the Royal Institution of Great Britain in London.

Other previously unseen objects include apparatus built by the nineteenth-century physicist John Tyndall, discoverer of the greenhouse effect. Tyndall used the equipment to investigate the theory of spontaneous generation, which proposes that living matter can develop from non-living matter.

Just 300 objects from the Royal Institution's collection of 6,000 scientific artefacts are currently on display. But on 18 August the institution received permission from Westminster Council to redevelop its central London home. More than half of the £13 million (US\$24 million) required for the work has already been found.



ROYAL INSTITUTION

contaminated with Creutzfeldt–Jakob disease, a BSE-like disease that occurs naturally in humans, the researchers say (A. C. F. Colchester and N. T. H. Colchester *Lancet* 366, 856–861; 2005).

NASA turns off gyroscope to prolong Hubble's life

In an effort to extend the life of the Hubble Space Telescope, engineers have shut down one of its three remaining gyroscopes.

The gyroscopes are a key part of the system that stabilizes the telescope and keeps it focused on celestial objects.

Hubble was originally designed to operate on three gyroscopes, with another three in reserve.

NASA managers say that Hubble will be able to operate with just two gyroscopes, although scheduling and making observations will be more complicated. By keeping the third gyroscope in reserve, it will be possible to prolong the telescope's life until mid-2008, scientists say, an eight-month extension over earlier predictions. They hope Hubble can be kept operational until shuttle astronauts can service it again — a trip that is unlikely to happen before 2007, if at all, because of delays in launching the shuttle fleet.

The rocky road to success

Tackling the legal and ethical minefield associated with human embryonic stem-cell research is not for the faint-hearted. **Erika Check** meets one man who is relishing the challenge.

Kevin Eggan has always been a climber. He grew up in the flat cornfields of Normal, Illinois, but every summer his father would take him mountaineering among the tall peaks of the Cascade Range in the Pacific Northwest. "Summer was nirvana in the mountains," Eggan says. "In the winter I would go back to school in Illinois, and I would just wait for the next summer to come along so I could go climbing again."

Eggan is now assistant professor of molecular and cellular biology at Harvard University. His passion for climbing may explain some of his success as a scientist, which his colleagues say is marked by two things: enthusiasm and a dogged determination to tackle the toughest problems in his field.

Eggan's next project could be his most challenging yet. This autumn he hopes to use cloning techniques on human cells to study conditions such as diabetes and Parkinson's disease — work that isn't being done at any other US university. But even as he prepares to begin these experiments, opponents are striving to ban them — and they are using some of Eggan's own findings to support their case.

Bespoke cells

Working with developmental biologist Doug Melton at the Harvard Stem Cell Institute, Eggan wants to generate embryonic stem cells that match individual patients. To do this he will use a technique called somatic cell nuclear transfer — more commonly referred to as cloning.

Eggan will take a sample of skin cells from a patient, and will extract the nuclei from individual cells. Each nucleus will be transferred into a human egg stripped of its own genetic material. In theory, the egg will then divide and develop into an embryo containing only the patient's DNA. And in the process, the egg will reprogramme the nucleus back to its earliest state, which means that cells from the embryo should have the potential to generate all of the tissues in the human body.

Each of these embryonic stem cells will carry the distinctive genetic quirks of the patient who donated the skin cells. By studying these stem cells, Eggan hopes to understand why the donor patients became ill, and what might be done to cure them.

Eggan's introduction to this technique came seven years ago, when he joined Rudolf Jaenisch's lab as a graduate student at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts. Jaenisch was

Scaling the heights: Kevin Eggan's work and hobbies are characterized by enthusiasm and determination.



starting up a mouse cloning project to study the genetic changes that make it possible for adult cells to be reprogrammed.

Eggan attacked this cloning project with zeal. "I thought it was the most exciting and coolest thing that I had ever seen," he says. But grim reality soon set in. Eggan shut himself in the mouse warehouse every day, seven days a week, for about a year. But he kept failing. He

could make embryos, but they simply refused to develop any further than a single cell.

What saved Eggan was his persistence — and his love of video games. He began to approach cloning the same way he had dealt with the classic arcade game *Super Mario Bros.* — by refusing to be beaten by a tricky stage or enemy. "There are people who really relish getting past that challenge, and they replay the game over and over and over and over again until they figure out exactly how to do it. That's the kind of person you need to be to succeed at nuclear transplantation," Eggan says. "And fortunately, that appealed to me." His determination eventually paid off and enabled Jaenisch's lab to publish a series of key papers

"Kevin has this attitude that science can be fun, and he doesn't hesitate to do an experiment because it's too hard." — Doug Melton

on cloning and its effects on cloned animals¹⁻³.

Melton, who was using human embryos left over from *in vitro* fertilization treatment to study disease, heard about Eggan's persistence from Jaenisch. He and Eggan met up and hit it off. "Kevin has this attitude that science can be fun," Melton says. "And he doesn't hesitate to do an experiment because it's too hard." The pair decided to start a project that would use cloning to produce new human embryonic stem-cell lines. The lines would be tailored to individual patients with diabetes and disorders of the brain and nervous system.

Private function

In 2002, the pair began the difficult task of securing funding and ethical approval for their project. Two separate measures bar the project from receiving funding from the US government. One, the Dickey Amendment, is passed by Congress every year and prevents scientists from using taxpayers' money to create or destroy human embryos for research, such as cloning. The second measure is a four-year-old executive order passed by President George W. Bush. It bans federal support for work on embryonic stem-cell lines created since 9 August 2001.

Melton is funded by the private Howard Hughes Medical Institute in Chevy Chase, Maryland, and last year, Harvard created the Harvard Stem Cell Institute, which will use money from other private donors to fund stem-cell research. With the money in the bank, Eggan and Melton are ready to begin their project. They are just waiting for a final approval from a Harvard ethics review board, which they hope to get this autumn.

In the three years it has taken to lay the groundwork for the cloning experiments, Eggan has tackled another challenge facing the stem-cell field: the shortage of human eggs. But in doing so he may have unwittingly played into the hands of those who want to ban work on human embryonic cells.

Cloning to produce stem cells requires human eggs. But few women want to endure the uncomfortable cycle of hormones and surgery that accompanies egg donation. In June, Eggan told a scientific meeting that he had developed a technique that may eventually solve this problem⁴. Eggan has created cells with properties similar to embryonic stem cells by fusing embryonic stem cells from existing lines together with skin cells. The embryonic stem cells seem to reprogramme the skin-cell nuclei just as eggs do.

Eggan calls his creations 'hybrids' — genetic remixes that act a lot like embryonic stem cells, but don't actually come from embryos. The paper describing the work appeared last month⁵ — just as the US Senate was gearing up to debate embryonic stem-cell research. And so, perhaps inevitably, Eggan's findings have taken on a life of their own.

In May, the House of Representatives passed a measure to overturn the Bush restrictions on



R. HELLMISS-PERALTA/HARVARD UNIV.

"No one expected the social and political maelstrom over cloning to crescendo to this point. But as I've been around since the beginning, I've been able to cope with the gradual incline." — Kevin Eggan

stem-cell research funding. The House bill would allow funding for research on stem-cell lines derived from embryos left over from *in vitro* fertilization procedures. The Senate is now considering passing that same bill — perhaps as early as this month.

The third way

But opponents of the work have seized on studies such as Eggan's as a possible third way out of the debate. They say that instead of overturning the Bush ban, the Senate should simply fund research that aims to make new stem-cell lines without creating embryos.

On 12 July, Leon Kass, a conservative bioethicist and chairman of the President's Council on Bioethics, lauded Eggan's work in the *Washington Post* on the morning of a crucial Senate hearing on stem cells. "Every public-spirited American should be encouraged by these findings. We should be hopeful that a technological solution to our moral dilemma might soon be found," he wrote. The publication of Eggan's paper only added fuel to the fire. On 23 August, the same newspaper quoted Senator Tom Coburn (Republican, Oklahoma), a physician who has led the opposition to embryonic stem-cell research: "All this is confirmation we will see breakthroughs without compromising ethical standards," he said. "We're not going to have to go that way if we can just be patient and fund the basic science."

Eggan says that these debates are misleading, because his work does not create substitutes for stem cells derived from embryos. The hybrids still contain two nuclei: one from a

skin cell and one from an embryonic stem cell. So they have an abnormally high amount of DNA, and Eggan needs to work out how to remove the embryonic stem cell's DNA. Eggan adds that he has only just begun working with the hybrids, so it is not clear what they will or won't be able to do. "It's frustrating," Eggan says, "because they're implying that our work is a solution, which it is not yet. These are ideas in their most nascent stages."

However the hybrid research develops, Eggan will continue with his planned cloning experiments using private funds. But the political debates over stem-cell research could still threaten this work. Some in the Senate are pushing for legal bans on any kind of cloning, which would outlaw Eggan's experiments. These laws seem unlikely to pass at the moment, but their supporters raise the issue year after year.

All the controversy could make a young biologist wary. After all, it's difficult enough to get a scientific career off the ground, even without the threat that the government might ban your research. But Eggan is used to the controversy; he began graduate school in 1998, the year after scientists in Scotland announced the birth of Dolly the sheep, the first cloned mammal. "No one really expected the social and political maelstrom over stem cells and cloning to crescendo to this point," says Eggan. "But as I've been around since the beginning of all of this, I've been able to cope with the gradual incline."

No stranger to inclines of any kind, Eggan still finds time to indulge his passion for climbing. Last November, he celebrated his father's 70th birthday by joining him on a little hike — a 6,000-metre jaunt up Mount Kilimanjaro. ■

Erika Check is Nature's Washington biomedical correspondent.

1. Rideout, W. M. III *et al. Nature Genet.* **24**, 109–110 (2000).
2. Humphreys, D. *et al. Science* **293**, 95–97 (2001).
3. Eggan, K. *et al. Nature* **428**, 44–49 (2004).
4. Check, E. *Nature* **436**, 9 (2005).
5. Cowan, C. A., Atienza, J., Melton, D. A. & Eggan, K. *Science* **309**, 1369–1373 (2005).

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

A wing and a prayer

Sightings of the ivory-billed woodpecker, a bird believed extinct for 50 years, have fired the public's imagination. But is it really alive? **Rex Dalton** joins the team trying to save this elusive bird.

The Big Woods in Arkansas is not a good place to be on a hot summer day. The swampy forest is thick with mud, poison ivy and snakes. Yet early last month, a dozen scientists slogged their way through these bottomlands towards a mesh tent abuzz with insects — the heart of an unusual US environmental project.

The group's goal is to save the ivory-billed woodpecker (*Campephilus principalis*), a magnificent bird thought to have died off at least 50 years ago as its forest habitat was chopped down. In April, a team led by ornithologists at Cornell University in Ithaca, New York, stunned the birding world by saying they had evidence that the woodpecker still lived in the Big Woods (J. W. Fitzpatrick *et al. Science* **308**, 1460–1462; 2005). Now more scientists are braving the wilderness as part of a federally sanctioned 'recovery team', charged with plotting a course to produce a healthy population of the birds.

Near the Cache River National Wildlife Refuge, where the ivory-billed was reportedly rediscovered in February 2004, the team has cut the trunks of trees to create deadwood. The deadwood, in turn, should become home to insect larvae, which are the woodpecker's favourite food. If the elusive ivory-billed shows up to snack, the scientists can use the insects in the tent to identify the larvae and better

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Some ornithologists fear that sightings of the ivory-billed woodpecker (top) have actually been of the lookalike pileated woodpecker (above).

understand the bird's eating habits.

To some, the team's mission is a conservationist's dream, a chance to bring back an iconic bird from the brink of extinction. To others, it is a near-futile attempt — possibly at the expense of other worthwhile projects — to save a species that may no longer grace these woods. The problem is that no one can prove whether the woodpecker still exists.

Forest hunch

"The bird is here," Martjan Lammertink insists in the steamy Arkansas bottomlands. Lammertink, a graduate student at the University of Amsterdam, has studied rare woodpeckers from Mexico to Cuba to Borneo. He moved to

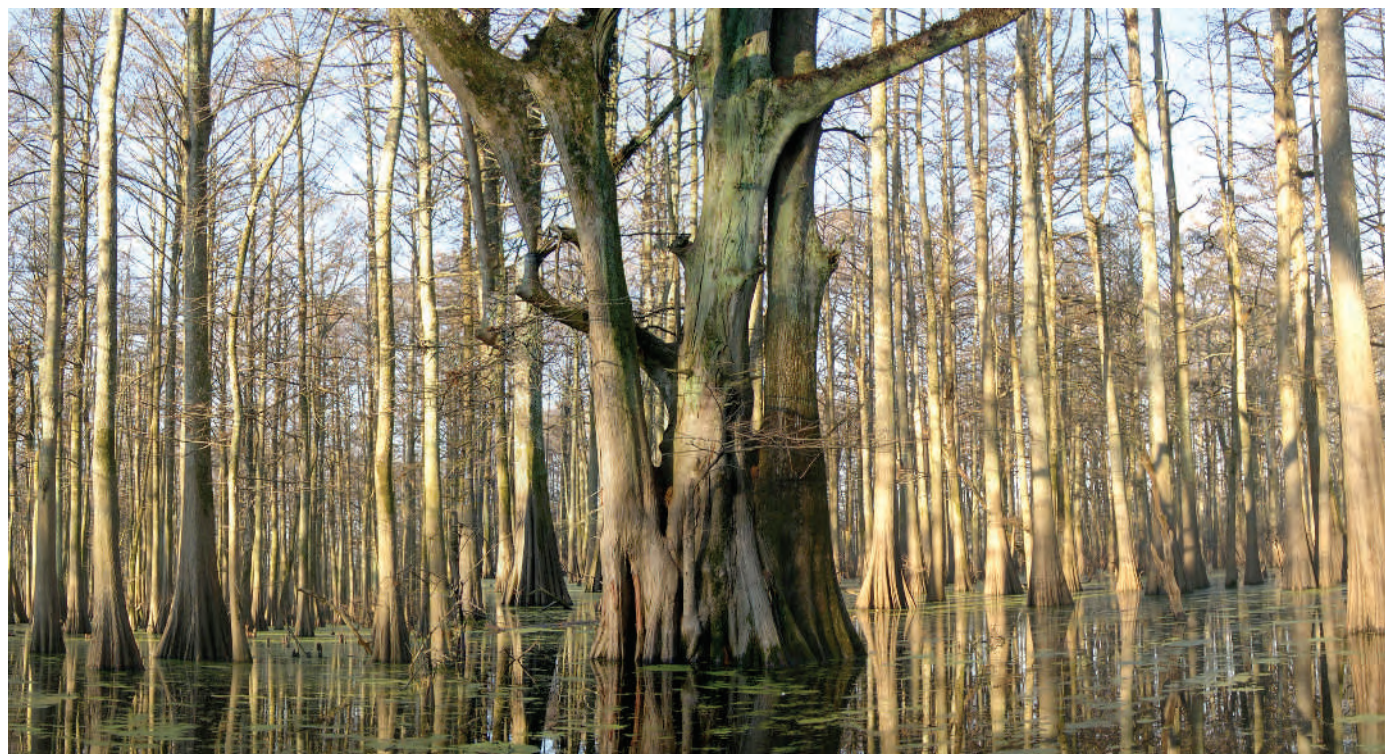
the nearby town of Brinkley so that he could devote himself to pursuing the 'Lord God bird' — named after what people would exclaim when they saw its gleaming bill, ferocious red crest, and body nearly half a metre long.

Yet not all ornithologists are as convinced as Lammertink. Several, including Jerome Jackson of Florida Gulf Coast University in Fort Myers, have repeatedly questioned the Cornell team's evidence for the bird's existence. Neither the four-second video purporting to show an ivory-billed, nor the sound recordings released last month at a bird meeting in California, satisfy this group of sceptics. Ivory-billed searchers, Jackson says, "are shooting in the dark".

The quest for the ivory-billed is steeped in politics as well as science. The Bush administration, smarting from criticism over its environmental policies, hailed the purported rediscovery as a rare piece of good news. Cornell team leader, John Fitzpatrick, had hoped that his occasional birding companion, First Lady Laura Bush, would make the announcement. But news leaks soon turned into a flood, and the secretaries of the interior and agriculture held the press conference instead — on 28 April, the day after the manuscript on the find had been accepted by *Science*.

Now, the federal government is providing an extra \$10 million to save the woodpecker's

J. SARTORE/NATL GEOGRAPHIC/GETTY IMAGES (MAIN PICTURE); D. ADAME/AP



C. JONES/CORNELL LAB. ORNITHOL.

Using a 'bug tent' (below), bird-watchers hope to find tasty treats to entice the ivory-billed woodpecker into the open in the Big Woods in Arkansas (above).

habitat, money redirected from other conservation projects even as congressional Republicans consider cutting back on the protections of the Endangered Species Act. Researchers on the ivory-billed's recovery team face a formidable challenge — figuring out how to save a bird that may already be extinct. Even if it still lives, there may be only a handful of individuals left, not enough to save the species as a whole.

US scientists have brought other birds back from the verge of extinction, including the whooping crane (*Grus americana*) and the California condor (*Gymnogyps californianus*). But never before has the species in question been almost completely invisible. The search for the ivory-billed, one scientist says, is like looking for "a moving needle in one hell of a haystack".

Search party

Still, the hunt goes on, and last month the recovery team met for the first time in Little Rock to explore possible research fronts. Some ornithologists are developing a 'life-table model' to try to determine how many of the birds could have survived, factoring in estimated lifespan and habitat. The best guess is about 15 pairs. Other scientists are examining the chisel-like mark of the ivory-billed's beak to see if it can be discerned from that of other woodpeckers. And major efforts are under way to increase the habitat for the bird's larvae hunt — the woodpecker can roam 20 kilometres a day in search of dying trees from which to strip bark.

Such studies are rarely controversial, but the \$10-million habitat-preservation plan has



raised some eyebrows. At the same time that biologists at the Fish and Wildlife Service (FWS) saw their budgets slashed, the ivory-billed project got a cash infusion. Meanwhile, successful programmes to recover other endangered species were scaled back.

One such example is the Kirtland's warbler (*Dendroica kirtlandii*), a small bird that breeds in Michigan and winters in the Bahamas. More than 30 years ago, the warbler population included some 160 singing males. Today, after years of trapping brown-headed cowbirds (*Molothrus ater*) that parasitically use the warblers' nests, the population boasts a record number of more than 1,400 singing males. But this spring, cowbird trapping in Michigan was drastically cut back; only about 1,100 cowbirds were snared compared with about 4,000 the previous year.

Federal biologists cannot prove that the warbler programme is suffering because of the

ivory-billed woodpecker, but many find the timing curious. "It doesn't make any sense to put one species at risk to save another," says Eric Carey, parks director at the Bahamas National Trust in Nassau, who studies the warbler's winter habitat.

FWS biologist Jon Andrew, who manages federal refuges in the southeastern United States, argues that such diversions are justified given the ivory-billed's precarious situation. "There was enough evidence that we ought to be acting as if the bird was there," he says. "When you make decisions on spending money, there are winners and losers."

To the public at least, the ivory-billed woodpecker is clearly a winner. The rediscovery announcement triggered a rush of national goodwill. Headlines trumpeted the bird that seemed to have made it against all odds. Arkansas experienced a miniature tourist boom as birders flocked to the refuge.

Elvis lives

In the face of this excitement it seems hard to believe that the FWS had been in the process of declaring the ivory-billed woodpecker extinct. No one had definitively seen one in the United States since 1944, although a separate population may have held on in Cuba until at least the 1980s. When the Cornell ornithologists finally thought they had spotted an ivory-billed, they were so amazed that they called the bird 'Elvis', after Elvis Presley, who was born nearby in the Mississippi River delta, and whose fans claimed to see him long after he died in 1977.

Yet, much as for the real Elvis, evidence for Elvis the woodpecker was greeted with

R. DALTON

scepticism. For years, rumours of ivory-billed sightings had swirled through the community but were never substantiated. In March 2003, a group of diehard ivory-billed seekers thought they heard and saw the bird in the White River National Wildlife Refuge, about 75 kilometres south of where Elvis would later be spotted.

Arizona naturalist Mary Scott posted a report of the sighting on her website, which birders monitor regularly. Among those drawn back to Arkansas was Tim Gallagher, a journalist who edits *Living Bird* magazine at Cornell. His subsequent trip and reported sighting in February 2004 launched the university's project with local birdwatchers.

Hard knocks

Over the next winter, when the trees were bare of their thick foliage, the team watched both refuges for months, using paid observers and their friends. They installed models of ivory-bills on trees, hoping to attract the real thing. They set out sound recorders to capture the woodpeckers calling or knocking on trees.

The effort echoed a similar quest 80 years earlier. By 1924, ivory-billed woodpeckers were supposed to be extinct, victim to rapid logging and to the fashion for collecting dead rare birds. But Cornell professor Arthur Augustus Allen, one of the world's leading ornithologists, photographed a pair of the birds that year in Florida. His rediscovery, which made headlines across the United States, triggered an intense effort to document the birds before they vanished for good.

Unlike Allen's group, the modern ivory-billed searchers had the benefit of videotapes and up-to-date recording techniques. After more than a year, the team had seven woodpecker sightings it considered believable, 18,000 hours of sound recordings, and one grainy video of Elvis.

When *Science* put the video on its website, it immediately provoked a sceptical reaction from some leading ornithologists. Three of them — including Jackson, who, despite his doubts, is a member of the recovery team — helped prepare a manuscript arguing that the bird in the film was probably a pileated woodpecker (*Dryocopus pileatus*), which shares similar red, black and white markings with the ivory-billed. The other co-authors on the paper were Richard Prum of Yale University in New Haven, Connecticut, and Mark Robbins of the University of Kansas at Lawrence (see *Nature* 436, 447; 2005).

On 1 August, as the journal *PLoS Biology* was close to publishing their critique of the *Science* article, Prum and Robbins withdrew the manuscript while Jackson was travelling. The Cornell team had offered them new sound recordings in support of the ivory-billed, including the bird's 'kent, kent' call and double-rap knock. The pair had heard other *Campephilus* woodpeckers in the jungles of South America make the same distinctive double-rap as on the new tape. And Cornell's



As the ivory-billed woodpecker's range dwindled (top), ornithologists in the 1930s turned to sound recordings to capture the bird's voice.

'kent, kent' call matched recordings of the bird made in 1935.

The evidence came from the modern remote recordings, only one-fifth of which have been fully examined. At a meeting of the American Ornithologists' Union in California in late August, Cornell scientists marshalled their evidence in several talks. Fitzpatrick's keynote lecture was greeted with thunderous applause.

Even so, the three sceptics say that they withdrew their *PLoS* manuscript too hastily. They are getting support from other ornithologists, including Gary Graves, the Smithsonian Institution's curator of birds, who argues that the bird shown in the crucial video may be a pileated, not an ivory-billed, woodpecker.

And some authorities say that there is unpublished evidence that helps prove that the bird in the video is a pileated woodpecker. They are referring to a 1953 film of a flying

imperial woodpecker (*Campephilus imperialis*), a species extinct in its home range of western Mexico. Some years ago, Lammertink secured a copy of the film, which had been taken by a birding enthusiast. It was among the evidence shown to ornithologist Michael Patten, research director at the University of Oklahoma's Sutton Avian Research Center in Bartlesville, when he visited Cornell in June.

Ornithologists generally agree that the imperial woodpecker is a sister bird to the ivory-billed, with many similar characteristics from coloration to the distinctive double-rap. But Patten was struck by the imperial's flight patterns. "As soon as I watched the film," he says, "I was absolutely certain they didn't have an ivory-billed woodpecker. The bird in the film flies utterly differently to the one in the Cornell video."

Sound basis?

Fitzpatrick is not troubled by the film of the imperial woodpecker, arguing that it sheds little light on whether his video shows an ivory-billed. "They are like apples and oranges," he says of the two videos, because of different camera angles and stages of the birds' flights.

To the sceptics, the strongest evidence to support the theory that Elvis lives are the sound recordings. And although Prum and Robbins are impressed by them, Jackson, Graves and Patten are more cautious. The evidence is tantalizing, they say, but not conclusive. "The sound recordings don't validate the flimsy sightings records," says Patten.

Remote sound recordings are notoriously deceptive, he points out. Blue jays (*Cyanocitta cristata*) can mimic woodpecker sounds, or gunshots can be mistaken for woodpecker raps. There are also untold copies of old ivory-billed recordings in private hands, including birders who play them back in the Arkansas refuges hoping to get a response from a live bird.

"We are 100% sure" that the new tapes don't just capture playbacks of old sounds, says Kenneth Rosenberg of the Cornell team. Analysis of the notes and sound shows that they don't match, he says. But the team acknowledges that it can't yet rule out a wily jay mimic.

For Fitzpatrick, all the evidence creates a critical mass that indicates at least one bird is out there. "There are people who need to see it to believe it," he told the ornithology meeting. "I respect that." But although he hasn't been fortunate enough to see one, he is "absolutely convinced" that Elvis lives.

Come November, Team Elvis will be back in the Arkansas woods with more observers, more acoustic recorders and an expected \$200,000 in federal funds. Already, Cornell and its partner the Nature Conservancy have raised more than \$5 million in pledges to fund the search and habitat preservation. If Elvis does indeed live, he is being offered one of the best chances ever for his species to survive. ■

Rex Dalton is *Nature's* US west coast correspondent.

"It doesn't make any sense to put one species at risk to save another."

— Eric Carey

BUSINESS

Fears rise over leaks of clinical trial results

Physicians in the United States are facing a fresh conflict-of-interest scandal over allegations that many of them accept cash for briefing financial analysts about the progress of clinical trials.

Thousands of doctors are paid to participate in teleconferences with stock-market analysts. And although most doctors say they are careful not to divulge confidential data, critics of such briefings say they wouldn't happen unless investors gleaned inside knowledge from them.

Senator Chuck Grassley (Republican, Iowa), chairman of the senate finance committee, last month demanded that both the Department of Justice and the Securities and Exchange Commission look into the practice, after a *Seattle Times* investigation cited 26 alleged examples of confidential leaks from ongoing drug trials.

Any investigation is likely to focus on hedge funds — loosely regulated financial instruments that allow investors to gamble on sharp movements of particular stocks.

Critics contend that physicians' responses to seemingly innocuous questions about drug performance serve as useful data for stock analysts. Professional bodies such as the Association of Clinical Research Professionals say they are watching the issue closely but have no authority to address it — although this might change if ethical standards are revised. PhRMA, the drug-industry organization, said in a statement: "Our laws are meant to protect patients, future research and the public. Breaking these laws is a reprehensible breach of trust and we condemn it."

Defenders of the briefings argue that talking to doctors directly is the best way for people issuing advice on drug-development investments to stay informed. In recent years, 'match-maker' companies have been created to facilitate such conversations. The Gerson Lehrman Group (GLG) based in New York is the largest matchmaker with some 60,000 physician clients.

Hourly fees that run from \$200 up to \$1,000 signal the importance placed on this information. Because pharmaceutical industries do not publish data from their own failed trials and independent researchers are often constrained by non-disclosure agreements, every bit of information provides a competitive edge for investors, says Charles Firreno, industry analyst with Life Science Insights, a market-research company based in Framingham,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Eric Topol (right) is opposed to the briefing of financial analysts by doctors involved in drug trials.

Massachusetts. Just one statement made by a physician can influence share price, he adds: "That situation is a clear conflict."

Investors can also use the information to sell stocks short, where trials indicate problems with a drug. Eric Topol, a cardiologist at the Cleveland Clinic in Ohio and former paid consultant, has publicly described investor practices used to obtain critical drug-trial information (E. J. Topol and D. Blumenthal *J. Am. Med. Assoc.* **293**, 2654–2657; 2005). The paper was published after Topol had been accused by *Forbes* magazine of providing insider information to a hedge fund which sold Merck stock short before problems with its painkiller Vioxx became public last October. "Managing the potential conflict is virtually impossible," says Topol.

The appearance of impropriety is enough for some to decline such consulting offers. Brian Druker, a Howard Hughes Medical Investigator at Oregon Health & Sciences University in Portland, left GLG two years ago on ethical grounds. It wasn't possible to be sure that he could always remember what was public, and what was confidential, he says.

"At its most egregious, this is a deliberate effort to unblind clinical trials for short-term profits," says Stanley Crooke, chief executive of Isis Pharmaceuticals, a company based in Carlsbad, California, whose stock dropped 20% in late 2002 after an industry report leaked problems with Isis leukaemia drug Affinitak. At the very least, it undermines the public's confidence in the drug approval process, Crooke says. ■

Virginia Gewin

IN BRIEF

NOVARTIS GRABS CHIRON

Swiss-based drug maker Novartis has offered to buy out California vaccine manufacturer Chiron for US\$4.5 billion.

The bid was made as the US Food and Drug Administration said it thought Chiron was making "significant progress" in addressing problems at its production plant in Liverpool, UK. Chiron stock has been depressed since UK regulators ordered it to cease production in Liverpool of its flu vaccine, Fluvirin, last October.

Analysts say that Novartis is moving to procure Chiron at a bargain price at a time when concerns about a global flu epidemic could boost the vaccine company's prospects.

Novartis already owns 42% of Chiron; it will buy the rest if its offer is accepted by most other shareholders.

CLOUD OVER SATELLITES Even as hurricane Katrina devastated the southern United States, industry officials admitted that a major weather satellite programme is running two years late and billions of dollars over budget.

Officials at Northrop Grumman in Los Angeles, the defence contractor that is building the satellites for the Department of Defense, told a conference in California that delays were due to problems with integrating many different types of sensor into the satellites.

The programme is now expected to be functional in 2010 at the earliest and to cost much more than the original estimate of \$4.5 billion.

NANOTECH IN MEDICINE Sales of medical products and equipment that use nanotechnology will grow strongly, a market research firm says — but perhaps not by enough to support the flood of activity in the sector.

Frost & Sullivan predicts that the world nanobiotechnology market — basically the use of nanoscale materials and devices for drug delivery, tissue reconstruction and other medical applications — will grow from US\$750 million last year to about \$2 billion by 2011.

But the company warns that too many players are joining the field and that regulatory obstacles could make it hard for nanotech-based products to compete.

Cuban science democratic and not tied to profit

SIR — We question the comparison made in your News Feature “¿Vive la revolución?” (*Nature* **436**, 322–324; 2005) between Cuban government-funded science and a corporate research programme using a top-down approach that focuses on applied science at the expense of basic research.

Reality contradicts the common view that science is undemocratic in Cuba — and that it is democratic in the United States.

In our experience as collaborators with Cuban scientists, science issues are first raised in local communities, then discussed at local research institutes and universities, then passed to higher levels of national ministries and the Congress, and there winnowed and prioritized. This is an up-and-down system that offers individual citizens a high level of engagement with decision-making processes for scientific research.

Neither have we found basic research to be neglected. For instance, there is work in underwater archaeology, palaeontology, plant and animal geography and many other topics. Indeed, Cuba probably has a better record in funding basic research than most other Latin American countries.

Both publicly and privately funded research in the United States and Europe, on the other hand, is often determined by corporate or political interests. Much research in the public interest withers for lack of resources, in favour of projects that will lead to patents and profits. Increasingly in the United States, the results of scientific research are also distorted or ignored by federal policy-makers if the science is inconsistent with the prevailing political agenda.

We should be more thoughtful about the Cuban system and our own.

Catherine Badgley*, Ivette Perfecto†

*Museum of Paleontology, University of Michigan,

†School of Natural Resources and Environment, University of Michigan, Ann Arbor, Michigan 48109, USA

Other signatories of this letter:

John Vandermeer Department of Ecology and Evolutionary Biology, University of Michigan

Gerald Smith Museum of Zoology, University of Michigan

Richard Levins Harvard School of Public Health

Bioweapons could kill more in one strike than guns

SIR — David Whitlock, in Correspondence (“Bioterror killed five in US; guns kill 30,000 a year” *Nature* **436**, 460; 2005), is right about the number of people killed each year by firearms in the United States. But in fact suicide accounts for more than half of these

deaths. The annual rate of firearm murder and non-negligent homicide is less than 10,000, and is overwhelmingly more often due to handguns than to military-style firearms (Federal Bureau of Investigation *Crime in the United States: Uniform Crime Report 2003*, Washington DC; 2004).

The firearm is designed for use against individual targets and is incapable of having the same large-scale effect as a well-dispersed biological agent. Almost 3,000 people perished in the 9/11 attacks, but a systematic biological release could conceivably claim 30,000 lives or more. We did not realistically anticipate commercial airliners being used as massively destructive devices, and are now in the difficult position of trying to determine how other major attacks could reasonably be prevented.

The firearm should not be considered in the same context as the potential nuclear, biological or chemical threat.

Michael C. Wendt

Washington University Medical School, 4444 Forest Park Boulevard, Box 8501, St Louis, Missouri 63108, USA.

EURYI scheme aims to stop women disappearing

SIR — Darach Watson and colleagues, in Correspondence (“Mysterious disappearance of female investigators” *Nature* **436**, 174; 2005), have pointed out that there were only three women among the 25 recipients of the first year’s European Young Investigator (EURYI) awards.

The first year of the EURYI award scheme has been externally evaluated, in a process that included sending questionnaires to 671 applicants, with a 70% response rate, and to all participating organizations (100% response). We also interviewed 20 people involved in the selection. Although these data could not be released while the evaluation was under way, the report is now publicly available at www.esf.org/euryi. Raw data may be made available on request.

As a result of the recommendations emerging from the evaluation, the European Science Foundation and organizations participating in EURYI have made a number of improvements to the second year’s processes, including ensuring that equal opportunities (EO) are provided to all applicants. Specific improvements include introducing at European-level selection an EO statement in refereeing/assessment documents and specific briefing on EO issues at the beginning of all peer-review meetings, such as interviews.

This approach is supported by results from the second year’s EURYI awards. Of the 672 initial applicants for the second year’s awards, 24% were women; 24% (31) of the

131 submitted to European selection were women, and of the 25 final recipients of awards, five (20%) were female.

The real issue for the European Science Foundation and EURYI participating organizations is raising the proportion of female applicants at the initial stages of this competition. To achieve this goal, we are adapting our publicity for the next call for EURYI proposals.

Neil Williams

European Science Foundation, 1 quai Lezay-Marnésia, BP 90015, 67080 Strasbourg cedex, France

EURYI: present procedure risks conflicts of interest

SIR — In their Correspondence “Mysterious disappearance of female investigators” (*Nature* **436**, 174; 2005), D. Watson and colleagues reported evidence of gender-based bias in the evaluation of applications for the first European Young Investigator (EURYI) awards. I agree with their analysis, but there were even more serious flaws in this process.

EURYI applications were first reviewed by participating national research funding organizations. About 83% of the proposals were rejected at this stage, the rest being sent for subsequent evaluation at European level. However, the initial unsuccessful applicants typically received only a very short rejection note, with no information about peer review.

The average annual grant size of EURYI is €200,000 (US\$239,000), which, in the case of my home country, Hungary, is several times more than the funding an established, productive researcher can apply for. This creates a major conflict of interest, which probably also holds true for other participating countries such as Austria, Denmark, Belgium, Ireland, Finland and Portugal.

I think the sharp drop in the number of applications for the second EURYI call in 2004 reflects the disillusionment felt by the participants.

A better procedure would have been for applicants first to submit short preliminary proposals, all of which would have been evaluated at the European level. The best applicants could then be asked to submit detailed documents for in-depth review. This would save a lot of time and effort for applicants and reviewers.

Gábor Lente

Department of Inorganic and Analytical Chemistry, University of Debrecen, H-4010 Debrecen, POB 21, Hungary

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

BOOKS & ARTS

In our own hands

Biotechnology is changing the world, but why do nations respond so differently?

Designs on Nature: Science and Democracy in Europe and the United States

by Sheila Jasanoff

Princeton University Press: 2005. 344 pp.
\$35, £22.95**Mark Cantley**

Biotechnology, like that valuable animal the scapegoat, is a beast destined, or chosen, to bear many burdens. But is this fate deserved? The perception — widespread in Europe — that biotechnology is something fundamentally new, like the discovery of electricity, or akin to black magic, is unfortunate. It has led to the assumption that there are technology-specific risks requiring *ad hoc* regulations and associated bureaucracies, and to consequent conflicts with sectoral regulations, as well as to international trade disputes. But not for the first time, perceptions, laws and the course of societal development may be driven by delusion. It was, after all, the scientists themselves who in 1974 proposed and accepted a moratorium on certain experiments, and the following year held the widely publicized international conference at Asilomar in California to discuss the conjectural risks of recombinant DNA.

Sheila Jasanoff has written a carefully structured, ambitious and timely book, *Designs on Nature*, about the evolution of public policy on biotechnology over the past three decades in the United States, Germany, Britain and the European Union (EU), and uses this as a basis for broader conclusions. Her central idea is correct: policy arguments about biotechnology have been hijacked to serve other agendas. Opponents of multinational corporations, of the industrialization of agriculture, of US policy and of globalization have found common cause with environment ministries that seek to enlarge their powers, and with non-governmental organizations that translate GMO as 'Greenpeace membership opportunity': they all have reasons to demonize biotechnology.

Scientists invited to public or government-sponsored forums on biotechnology soon found that the skills they needed were more akin to mud-wrestling; their expertise and experience were considered disqualifications. As in her earlier work, Jasanoff disputes that the supposedly neutral, objective, value-free process called science can be separated from the social and political matrices in which it is embedded. Rather, she emphasizes "the

constructed and value-laden character of scientific knowledge", and asks whether science will lose its ability to serve either state or society as a source of impartial critical authority.

Jasanoff has read widely, has broad awareness and has spoken to many people in the countries she discusses. She marshals her information carefully, using a comparative approach to illustrate how similar challenges to public policy-makers in these countries were handled differently, in ways that reflect long-standing differences in their political cultures. She focuses on "civic epistemology, the institutionalized practices by which members of a given society test and deploy knowledge claims used as a basis for making collective choices".

The book has weaknesses, however. The author, a lawyer and policy analyst rather than a scientist or historian, makes numerous errors, not all of them minor. Crick and Watson were not working at Cambridge University, and their Nobel was not for "cracking the genetic code". Legislation such as the EU directive on biotechnology patents, which Jasanoff discusses at length, was adopted by co-decision of the European Parliament and the Council of Ministers — not the European Commission, whose constitutional position in such legislation is restricted to drafting the proposals.

Jasanoff's use of historical evidence appears selective, and some of the omissions are eloquent counter-testimony to her thesis. Asser-

tions about the separateness of the regulatory debates in the three countries (and the EU) are contradicted by several years of expert debate at the Organisation for Economic Co-operation and Development (OECD), which culminated in the much-cited report *Recombinant DNA Safety Considerations* (OECD, Paris, 1986) and unanimous Council Recommendation recognizing "that there is no scientific basis for specific legislation to regulate the use of recombinant DNA organisms". In the late 1970s and early 1980s, there was close bilateral dialogue between key players in the US administration and the European Commission, with direct consequences for the 1982 Council Recommendation on recombinant DNA. Similarly, she discusses the concept of "substantial equivalence" in the context of genetically modified food regulation, but again without reference to the OECD debate and report *Safety Evaluation of Foods derived by Modern Biotechnology* (OECD, Paris, 1993), which popularized the use of this concept.

The author's perspective on societal learning processes fails to acknowledge their dynamic character. She notes that "as biotechnology inexorably moved from the cloistered scientific laboratory to the competitive marketplace, it became clear that the post-Asilomar settlement could not continue unamended". Yet she describes as an "about-face" what others might say were rational responses to the

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Hostile reception: some people have demonized biotechnology, particularly genetic modification.

progress of knowledge and the accumulation of experience.

One might dispute Jasanoff's assertion that there "was no storehouse of precedents that policymakers could reach into for historically documented evidence concerning the widespread use of laboratory-crafted organisms". Modern biotechnology brings precision, but no fundamental change, to the long-standing practices of selective breeding and random mutagenesis, which had been used in the fermentation and seed industries for many decades before the debates on recombinant DNA.

Jasanoff's thesis rests on stronger ground when she turns to the ethical impacts of modern biotechnology and genomics, and the differing national responses, which she discusses and compares in detail: "Genetic engineering transgresses some of the most deeply entrenched categories of western thought... Designs on nature — once thought to be the prerogative only of a divine creator — seem now well within the reach of human capability," she writes. The "controlling narratives" that framed the course of policy development include not only a novel process for intervening in nature, and a source of new products, but "a state-sponsored program of standardization and control carrying profound implications for human dignity and freedom, and raising questions of constitutional significance". Given the profundity of the challenges thus brought into public and policy debates, democratic theory in the era of the knowledge society must take on board the involvement of citizens in the production, use and interpretation of knowledge for public purposes.

She offers three main conclusions. First, that core concepts of democracy such as citizenship and accountability cannot be satisfactorily understood without considering the politics of science and technology. Second, that in all three countries (and the EU), policies for the life sciences have been incorporated into 'nation-building' projects that seek to reimagine what the nation (or Europe) stands for. Third, that political culture matters to democratic culture, and works through the institutionalized ways in which citizens understand and evaluate public knowledge. These three aspects of contemporary politics help account not only for policy divergences between nations, but also for the perceived legitimacy of state actions.

These conclusions are well supported, and useful not least for indicating why scientific, industrial or other communities have found it difficult to influence policy. So the book succeeds in its aims. The policy debates have served to crystallize the emergence of a European polity and self-awareness. But the reader is left wondering: what if these high policy debates are founded on misperceptions, which they reinforce? ■

Mark Cantley is adviser, Research Directorate-General of the European Commission, on Biotechnology, Agriculture and Food.

A scientist's life in Russia

About Science, Myself and Others

by V. L. Ginzburg,

Institute of Physics/Taylor & Francis: 2004.

549 pp. \$99.95

Alan L. Mackay

Vitali Ginzburg, winner of the 2003 Nobel Prize for physics, was born in Moscow in 1916. He has survived, through intellect, character and chance, through the whole tempestuous and tragic period of the Soviet Union. He experienced two world wars, a revolution, leninism, stalinism, the fall of the Soviet Union, and the chaotic present-day reconstruction of Russia at the hands of 'gangster capitalism'.

About Science, Myself and Others deals with some immense topics: the circumstances of the Soviet Union, the history of branches of fundamental physics, and the lives of distinguished physicists, especially Lev Landau. In particular, Ginzburg focuses on struggles in three main areas: fundamental physics, the mechanics of daily life, and the ideologies of politics and religion.

Since the rise of Mikhail Gorbachev in 1985, Ginzburg has produced various accounts of his life and work, biographical sketches of his contemporaries, and more general pieces. *About Science, Myself and Others* is an English translation of the latest Russian edition, from 2003, which brought together scattered material but with some repetition and no index. It also contains material based on chapters published in his earlier book *The Physics of a Lifetime* (Springer, 2001), and includes its contents list. The new collection contains a mass of detailed information on physics and on people, making it indispensable, particularly to those interested in the school of Landau and his fellow Nobel laureate Igor Tamm.

Ginzburg writes his *apologia pro vita sua* not as an apology but with a (justified) "fair conceit of himself" for posterity as his "version of the facts". He makes little concession to the ignorant and suggests a simple test of general knowledge for politicians and others. For example: "Q. What causes the seasons? A. The inclination of the axis of the Earth to the plane of its orbit." He writes that he has found interviews unsatisfactory, and in one paper he even wrote both the questions and his answers.

Most people have sought "to render unto Caesar what is Caesar's and to God what is God's", with a boundary between their inner beliefs and the society in which they live. But Ginzburg has always had an intense concern not just for physics but for the welfare of Russia. In 1989, during the Gorbachev era, he even became a deputy in the Supreme Soviet (the parliament), representing the Soviet Academy of Sciences, but left politics in 1991. In recent years he has been "an incorrigible

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Vitali Ginzburg looks back on the science that flourished amid the political turbulence in Russia.

atheist", and campaigned against the revival of the Orthodox Church and pseudoscience of all kinds.

Ginzburg never suffered personal repression from the authorities, and in fact was honoured by the state, primarily for his work on the nuclear-weapons programme. He made a key contribution to the physics of the hydrogen bomb, proposing the use of lithium-6 deuteride as a nuclear source — a suggestion that surprised the Americans and brought him, in late 1953, the Order of Lenin and the Stalin prize first class. He might otherwise have suffered repression from the state following his marriage to Nina Ermakova, who had been exiled to Gorki and whose father was falsely regarded as an enemy of the state.

If history is a line with branch points at arbitrary decisions, then the whole manifold of all possible trajectories might be viewed as a tree. If this is the history of Russia, what shape is this tree of all possible paths? How different could it have been? The science theoretician Arnosht Kolman entitled his autobiography of disillusionment *Our Lives Should Have Been Different* (1982). How would Nikolai Bukharin, for example, have dealt with the situation if he, instead of Stalin, had succeeded Lenin?

Karl Marx hoped to discover laws and regularities in human history, as did the 'general systems' movement from Bertalanffy to Peter Turchin today, but the practical results have been slight. Ginzburg points to the key branch point: "I believe that the fundamental and principal cause of all these disasters is the Bolshevik-Communist totalitarian regime which was set up in Russia as a result of the *coup d'état* in October 1917," which followed the February 1917 revolution that deposed the tsar.

The first part of the book is a detailed review of the physics that concerned Ginzburg, including radiation from uniformly moving sources (the Vavilov-Cherenkov effect), cosmic rays, soft modes, superconductivity and superfluidity. It manages to place his own

work in a historical framework.

The second part of the book deals with history and includes biographical memoirs of several of his contemporaries, notably Landau, Tamm and Evgeni Lifshitz, but also Sergei Ivanovich Vavilov (brother of the geneticist Nikolai Vavilov and former president of the Soviet Academy of Sciences), as well as an autobiography. There is also an analysis of Russia's relationship with the Nobel prize, particularly in connection with C. V. Raman's prize for Raman scattering and the parallel

work of G. S. Landsberg and L. I. Mandelstam, and with the discovery of the Vavilov-Cherenkov effect. Finally, there is a collection of Ginzburg's publications on social and political questions, written with great factual detail and strong feeling.

In short, Ginzburg presents himself and his work, for the record, in the wider world of physics and of Russia. ■

Alan L. Mackay is in the School of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK.

the ethical and philosophical implications of our own biology is not taken. Towards the end of the book, however, Forbes hints at a moral landscape. Somewhat negative language is used when describing current medical practices in assisted human reproduction. For example, the artificial implantation of multiple embryos increases the chance of a successful pregnancy, but inevitably it also increases multiple births of twins and triplets. This in turn creates its own problems (some as yet unknown) for the mother and especially for the future health and mental functioning of children born from artificial multiple pregnancies. The tragedy is that individual couples cannot afford the possibly healthier alternative of having fewer embryos implanted each time (because of the many more cycles that this would require), and most governments do not consider it their financial responsibility to step in, especially where older couples are concerned.

The analogy of Aldous Huxley's *Brave New World* is wheeled out to provide the usual warnings against the application of modern technology to human reproduction. Yet, evolutionarily speaking, Huxley's was perhaps a group-selectionist vision of excessive state control. The current 'customer-centred' health care that arises from free-market economics, and generates the social dilemma outlined above, might rather be described as an outcome of individual-level selection. But what, if any, should society's role be in moderating individual reproductive choices? I can't help feeling that the author, having implicitly introduced moral statements concerning modern human reproduction, bears some responsibility to discuss the ethical issues that follow from an evolutionary perspective — if only to expose the complexity of the problem.

At the end of the day, this foray into popular science perhaps does not do justice to the high quality of the author's own research on avian family conflict. In terms of outlining the underlying theory and non-human evidence, other texts, such as Douglas W. Mock's *More Than Kin and Less Than Kind* (Belknap Press, 2004), probably represent a better compromise between the popular and formal science formats. However, these do not attempt to deal explicitly with the tricky issues of human families and reproduction. In contrast, Forbes has taken on one of the hardest tasks in popularizing science, that of explaining to readers complex scientific findings about their own life and culture — and inevitably we all have our own opinions about that. ■

Jonathan Wright is in the Institute of Biology, Norwegian University for Science and Technology, Trondheim 7491, Norway.

Families behaving badly

A Natural History of Families

by Scott Forbes

Princeton University Press: 2005. 248 pp.
\$27.95, £17.95

Jonathan Wright

I always feel that I am in a better position than many scientists when a friend or family member asks what my research is actually about. Compared with a theoretical physicist, it is relatively easy for me, as a behavioural ecologist, to enlighten and entertain members of the public, using the obvious parallels that exist between birds and humans in their family relationships. For example, there are disagreements between members of a pair over who does most of the childcare, and constant battles between overworked parents and their demanding offspring concerning apparently unimportant items of food. In his new book *A Natural History of Families*, Scott Forbes takes up this theme with a vengeance to reveal the many fascinating aspects of human family life and reproduction that arise from evolutionary conflicts of interest.

This 'popular' science book sells itself on the salacious revelation that all is not as nice as it seems within our cosy families and in the wonderful creation of an infant. As an avian biologist, the author presents a dispassionate account of parent-offspring conflict, siblicide and infanticide in animals and plants. But in the background lurks the disturbing fact that we are talking here about mothers and fathers, sons and daughters, and the family — an institution at the heart of human social experience.

The full horror is revealed when the author exposes at length the human side of the story. He describes research showing that 'morning sickness' and even pre-eclampsia (dangerously high blood pressure) during human pregnancy may be the manifestation of an evolutionary struggle between mother and child over placental blood flow and the frequent abortion of 'unfit' embryos. Fathers also get involved, and often on the side of the offspring. Certain genes become 'imprinted' only when inherited from the father, and these prompt the embryo to sequester even higher

levels of resources from the mother.

Siblings are no better to one another, given the disturbing fact that some of us must have been murderers while in the womb: in more than 90% of human twin conceptions, mortality of one of the two embryos occurs. If we were a species of bird, this fact would classify us as an 'obligate brood reducer', along with those famously bloodthirsty little black eagle chicks.

How, then, do we react when we are told that these awful things are going on in our own families and inside our own bodies? More importantly, what do we think about the social and medical policies that perpetuate, or seek to modify, these phenomena? Does evolution provide a rationale for their continued existence, or a starting point for a better-informed programme of cultural and political change?

Unfortunately, the opportunity to explore

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Who said we are meant to live in harmony? Malcolm in the Middle sees the funny side of life in a dysfunctional family.

STRUCTURAL BIOLOGY

Fibres hinge on swapped domains

Andrew D. Miranker

When proteins assemble themselves into fibres, there can be grave pathological consequences. Designing an otherwise soluble protein to make fibres provides a general mechanism for the construction process.

DNA encodes RNA, which encodes protein — this ‘central dogma’, coined by Francis Crick in 1958, underpins modern molecular biology¹. Implicit in this formula is the tenet that a linear chain of amino acids contains the complete instructions to form a three-dimensional protein structure. Despite an astronomical number of possible configurations, a functional protein independently adopts the structure with the lowest free energy². These axioms are so universally observed that their violation holds particular fascination. A functional example of this is α -lytic protease, whose unfolded state is more stable than its native conformation³. A pathological example of where the rules appear to be broken is when proteins aggregate into a form that is more stable than their functional precursors — a process found in a host of diseases, such as Alzheimer’s, Huntington’s and prion diseases⁴. In this issue, Sambashivan *et al.* (page 266)⁵ reconcile this apparent contradiction between dogma and observation.

The fact that normally soluble proteins are capable of aggregating comes as no surprise to anyone who has worked with them — finding a solid white mass at the bottom of your test tube is a well-known frustration of the field. Careful analysis of this phenomenon, however, reveals that in many cases the aggregates are structured, linear assemblies typically called amyloid fibres. Regardless of the starting conformation of the precursor, amyloid fibres are always formed about a central core whose polypeptide chain is organized from strands that stack at right angles to the long axis of the fibre. The initial formation of such structures is a rare event. Once present, however, the fibres form templates that catalyse their own formation. The resultant structures are exceptionally resistant to degradation and disassembly by chemical or proteolytic means. But hang on — a protein structure that requires a protein template? Alternative structures with greater stability than the

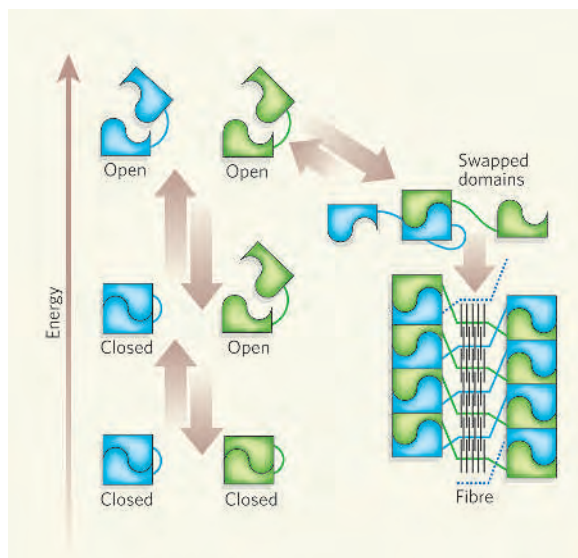


Figure 1 | Structures and energies associated with assembly of domain-swapped proteins. Some proteins adopt a structure characterized by two well-defined domains tethered together by a loop. If the loop can serve as a hinge, the protein may open, transiently adopting a high-energy state. Very rarely, two such open states encounter one another. This can result in the formation of an open swapped dimer where the interface of the two domains is regenerated, but by two separate polypeptides. Sambashivan *et al.*⁵ engineered the hinge to include a short sequence that interacts with itself. The result is synergy between runaway domain swapping and the formation of an amyloid fibre backbone. (Adapted from Fig. 1c of ref. 5.)

native state? That goes against the dogma!

The proteins that behave like this are typically divided into two categories: those that assemble from globular precursors, such as β_2 -microglobulin which deposits as amyloid fibres in patients treated with dialysis; and those that assemble from peptide precursors, such as amyloid- β peptide, a peptide of 40–42 amino acids that forms fibres in the plaques of Alzheimer’s patients. In cases of the former, amyloid formation appears to result from the assembly of states that have both amyloid- and native-like features⁶. This suggests a role for native structure in amyloid assembly. In cases of the latter, the intrinsic disorder of the short flexible chain yields to the formation of a well-defined structure upon aggregation.

Sambashivan *et al.*⁵ reconcile these two

categories of amyloid formation with a mechanism that does not violate the protein-folding dogma. The proposal is based on three-dimensional domain swapping⁷. In this mechanism, a single polypeptide chain can be thought of as having two independently folded sub-domains bound together through complementary surfaces (Fig. 1). The two domains are separated by a portion of the protein that acts as a hinge. Occasionally, the hinge will open and close, transiently exposing the interface. The interface can also be reconstituted through collision with a second protein that, fortuitously, is also in the open state. This is a rare event as it requires the collision of two high-energy, and therefore poorly populated, states. If the dimer forms in such a way that only one of the two open interfaces is closed, a state arises with sticky ends. This is the nucleating structure, because it always presents an open overhang to bind to other polypeptides. Furthermore, this open domain-swap is a structure-specific template of its own extension.

The problem with domain swapping as an explanation for fibre formation is that it is energetically unfavourable. A collection of proteins will have more entropy if they are free to move and tumble independently. Swapped assemblies lose this entropy in exchange for an interface that was already present in the starting structure. Therefore, there must be something about a swap that makes it more favourable to be joined up than to be free and single. The key is in the hinge. One possibility is that the hinge itself can form interactions with other hinges. Sambashivan *et al.* investigated this using hinges that can interact with each other, and introduced them into a protein that is capable of swapping domains.

Because of their large size and variable length, amyloid fibres confound traditional methods used to determine protein structure. Sambashivan *et al.*⁵ use an ingenious approach to show that the protein in the fibre has a native structure. The design capitalizes on the

intrinsic ability of the enzyme RNase to form domain-swapped dimers. The active site of this enzyme includes two amino acids that are essential for catalysis and that reside on opposing domains. Mutating these amino acids separately generated two pools of inactive mutants. Upon domain swapping, these formed a pool of dimers, one in four of which is active — representing the probability of assembly of complementary mutants. The authors then engineered the hinge region to include the insertion of ten glutamine residues. Alone, this is a weakly amyloidogenic sequence; in the proteins that are mutated in polyglutamine-expansion diseases, such as Huntington's disease, stretches of more than 40 glutamines are required for significant aggregation⁸. Remarkably, a mixed assembly that includes the active-site and hinge-insertion mutations results in the formation of amyloid fibres with an enzymatic activity that is indicative of domain-swapped assembly.

Enzymatic activity is the unquestioned hallmark of a defined protein structure. In the founding days of protein crystallography, serious scepticism existed concerning the relevance of crystal structures to biological

function. Then, as now, RNase played a pivotal role in a seminal study⁹ that established that proteins in crystals can be enzymatically active. There is considerably more mystery to be solved in the formation of amyloid fibres — for example, the structural nature of transient intermediates and the basis of their cytotoxicity. The current study⁵ shows us that a molecular basis for these phenomena, with creative perseverance, is ultimately obtainable. ■

Andrew D. Miranker is in the Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06520-8114, USA. e-mail: Andrew.Miranker@yale.edu

1. Crick, F. H. C. *Symp. Soc. Exp. Biol.* **12**, 138–163 (1958).
2. Haber, E. & Anfinsen, C. B. *J. Biol. Chem.* **237**, 1839–1844 (1962).
3. Sohl, J. L. *et al.* *Nature* **395**, 817–819 (1998).
4. Dobson, C. M. *Nature* **426**, 884–890 (2003).
5. Sambashivan, S., Liu, Y., Sawaya, M. R., Gingery, M. & Eisenberg, D. *Nature* **437**, 266–269 (2005).
6. Eakin, C. M., Attenello, F. J., Morgan, C. J. & Miranker, A. D. *Biochemistry* **43**, 7808–7815 (2004).
7. Liu, Y. & Eisenberg, D. *Protein Sci.* **11**, 1285–1299 (2002).
8. Scherzinger, E. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 4604–4609 (1999).
9. Doscher, M. S. & Richards, F. M. *J. Biol. Chem.* **238**, 2399–2406 (1963).

BIOLOGICAL PHYSICS

Rare returns on lost effort

Wesley P. Wong and Evan Evans

How does the size of a system affect its thermodynamic irreversibility? A deft experiment that observes the unfolding and refolding of a single molecule of RNA provides insights into the question at a small scale.

When vigorously stirred, a cup of coffee will heat up; a cup of hot coffee, however, will never stir a spoon (Fig. 1). This type of irreversibility is a cornerstone of macroscopic thermodynamics. But does it still hold when the spoon is extremely small? Newly discovered relationships between heat, work and energy have revealed unexpected features of small systems when they are driven far from equilibrium — relationships that are becoming increasingly relevant as experiments probe ever-tinier systems, including the nanoscale machinery of living cells. On page 231 of this issue, Collin *et al.*¹ describe a single-molecule experiment that not only validates one prominent recent postulate, known as the Crooks fluctuation theorem, but also provides a new method for quantifying the difference in equilibrium free energy — the useful work expected to be extracted from a system — between two biomolecular states.

The Crooks fluctuation theorem² (CFT) describes the exchange of energy between a system and its environment in forward and reverse processes. It generalizes the so-called Jarzynski equality³, which relates equilibrium

free energy to the work done in multiple, non-equilibrium measurements. The application of the CFT to a single-molecule experiment can be understood by considering an idealized 'dashpot' — a plunger that can move through a viscous fluid held at a constant temperature.

The force required to move a large, slow-moving plunger is proportional to its speed; all the mechanical work performed is lost, irretrievably, to the environment as heat. At a given speed, the work required to move the plunger between two well-defined states is always the same: viewed microscopically, the resistance to the motion is just the statistically averaged effect of many collisions between the plunger and the fast-moving molecules of the fluid.

If the plunger is very small, however, this average is performed over far fewer collisions, and the random, brownian movements of the plunger become significant. When an excess of molecular collisions occurs in the direction in which the plunger is moving, the work done by the operator is reduced. This work can on rare occasions be reduced to zero, or even become negative — meaning that the system performs work on the operator. The mechanical work needed to move a small plunger is no longer a precise, deterministic variable; rather, it is described by a statistical distribution.

The CFT stipulates that the probability of exerting a given amount of mechanical work on the plunger in the conventional, forward process, divided by the probability of getting the same amount of work back from the reverse process, depends exponentially on the amount of work lost to the environment. This implies that the likelihood of dissipating virtually no work is the same in both the forward and reverse directions. Because the amount of work dissipated increases as a system gets larger, in a sufficiently small system the likelihood that 'dissipation-free' events will occur becomes large enough for them to be observed directly.

Collin *et al.*¹ studied the mechanical unfolding and refolding of individual RNA molecules, applying and measuring forces using optical tweezers — a focused laser beam that acts as an ultra-sensitive spring to trap and hold small particles. The two ends of an RNA molecule, folded into junctions and hairpin shapes, were each linked to a small bead, one held by a micropipette and the other by the optical tweezers. In each experimental cycle,



Figure 1 | Stirring stuff. A paradigm of thermodynamic irreversibility: the world's largest cappuccino.

S. RELLANDINI/REUTERS

the beads were moved apart to produce an unzipping or unfolding transition, and then moved back together to allow refolding.

By integrating the force over the change in length of the molecule through many cycles, Collin and colleagues¹ obtained separate statistical distributions for the work done on the molecule during the forward and reverse transitions. They used these to verify that the CFT holds. By finding the energy at which the two distributions crossed for a range of speeds of bead movement, the authors were able to evaluate the free energy of transition between the folded and unfolded state of an individual RNA hairpin. They also demonstrated the robustness of the CFT by using their analysis to quantify the equilibrium free energy accurately even when the beads were moved apart too fast for the system to respond — in other words, when the system was taken far from equilibrium.

Collin and colleagues' work clearly verifies that the CFT can be applied to small biomolecular systems. But what are its wider implications for single-molecule research? The method provides a unique approach to quantifying the free energy in two-state systems — even in those systems far from equilibrium, where the transition between the two states requires very different amounts of work depending on its direction (a phenomenon known as hysteresis). This is important, as it is often the drifting baseline of a measuring instrument that limits how slowly, and thus how close to equilibrium, an experiment can be performed.

The CFT method is unlikely, however, to be suitable for all molecular systems and probe techniques. Its success in Collin and colleagues' experiments is due in part to certain features of the RNA hairpins and junctions used that made force a relatively smooth function of displacement between the well-defined initial (folded) and final (unfolded) states. Specifically, the free-energy landscape between these states resembles a 'staircase' of small energetic steps of short duration spread over an easily observable distance of many nanometres. The nonlinear, elastic behaviour of the polymer linkers and unfolded RNA chain also ensured that forward and reverse responses could be easily aligned in the presence of instrument drift. These attributes, together with the exquisite sensitivity of the optical tweezers — in regard to both the force applied and the displacement caused — allowed accurate integration for calculating the work during both unfolding and refolding. It also allowed for the subtraction of energetic contributions from the optical tweezers and chemical linkers, a necessary step in precisely determining the free-energy difference of the molecular transition. Extending Collin and colleagues' technique to other biomolecular transitions is thus sure to present fresh challenges.

Even so, this exciting experiment has provided insight into the way that recently developed fluctuation theories can be applied

to transitions in biomolecular systems. Not only does physics teach us new ways to understand single-molecule measurements, but such measurements also feed back into theory, extending our knowledge of thermodynamics to systems that are small and out of equilibrium. Stir a cup of coffee with a small enough spoon, and the coffee might just stir you. ■

Wesley P. Wong and Evan Evans are in the Department of Biomedical Engineering, Boston

University, 44 Cummington Street, Boston, Massachusetts 02215, USA. Wesley P. Wong is also in the Department of Physics, Harvard University, 17 Oxford Street, Cambridge, Massachusetts 02138, USA.
e-mails: wesley@physics.harvard.edu; evanse@bu.edu

1. Collin, D. *et al.* *Nature* **437**, 231–234 (2005).
2. Crooks, G. E. *Phys. Rev. E* **60**, 2721–2726 (1999).
3. Jarzynski, C. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).

EVOLUTIONARY GENETICS

Microarrays and species origins

Roger Butlin and Cally Roper

Whole-genome arrays have been used to reveal small islands of genetic differentiation in *Anopheles* mosquitoes. Analysis of these regions will identify genes involved in the initial stages of speciation.

What are the genetic modifications that underlie the first steps in the origin of new species? The use of microarrays to make whole-genome comparisons between populations that can interbreed, but only to a limited degree, provides a fresh approach to the question. As reported in *PLoS Biology*, Turner *et al.*¹ have applied this methodology to the malaria-carrying mosquito, *Anopheles gambiae*. They have identified three small regions, less than 1.5% of the genome, that contain critical genes for the initiation of speciation.

In the classic model of speciation, two populations diverge during a period of geographical separation and cannot exchange genes when the physical barrier is removed. In this case the two genomes accumulate differences uniformly. There is, however, often gene flow during the period of divergence, or after secondary contact if reproductive isolation is incomplete. In these cases, genetic differentiation becomes unevenly distributed across the genome as gene flow erodes differences between species, leaving islands of differentiated sequence only around the genes responsible for reproductive isolation². This pattern has been nicely illustrated by the sequencing of multiple genetic loci in closely related species pairs of the fruitfly *Drosophila*^{3,4}, and by surveying differentiation at many anonymous markers, for example in the winkle, *Littorina saxatilis*⁵.

Whole-genome microarrays are used primarily for studies of gene expression, but they can also be used to detect differences in DNA sequences. Genomic DNA is hybridized to the array and a weak signal indicates a mismatch between the individual's DNA and the probe sequence. This principle has been successfully applied to the genetic mapping of quantitative traits in a plant, *Arabidopsis*⁶. Now, Turner and colleagues have used it to scan the whole mosquito genome for genetic

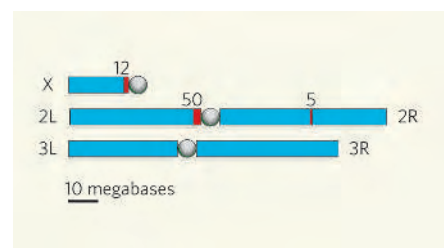


Figure 1 | Differentiated regions in the genomes of the M and S forms of *Anopheles gambiae*.

Anopheles gambiae has three chromosome pairs (X/Y, and 2 and 3, with the two arms of 2 and 3 being designated L and R). Numbers above the differentiated regions, shown in red, are counts of predicted genes within them: among these are genes critical for the initiation of speciation. The small differentiated region on 2R is least strongly supported statistically, perhaps simply because it is small. It may be significant that two of the differentiated regions are close to centromeres (circles).

variants that distinguish two forms of *Anopheles gambiae* known as M and S.

The M and S forms probably began to diverge very recently, perhaps within the past 10,000 years. The M form is restricted to West Africa, where it occurs alongside the S form, and the two interbreed at a low frequency (about 1.2%)⁷. The resulting hybrids are viable and fertile. Fixed differences unique to each species occur in the genes for ribosomal RNAs on the X chromosome⁸. But variable genetic markers elsewhere in the genome show little or no differentiation⁹.

Turner and colleagues' approach involved hybridizing DNA from seven individuals of each form, from Cameroon, to arrays of 142,065 probes each 25 base pairs long. Regions containing concentrations of differences between M and S were detected using a 300-probe sliding window, and just three small regions turned out to be significantly

differentiated (Fig. 1). On its own, this analysis may not have the power to distinguish regions protected from gene exchange from those that have large differences due to the chance effects of genetic drift. But sequencing of a sample of loci within the differentiated islands, close to these islands, or elsewhere in the genome, confirmed the result for a larger sample of individuals by detecting fixed nucleotide substitutions only in the candidate regions.

Among the small number of genes contained within these genomic islands of differentiation there must be critical genes for the first stages of speciation. The few currently known examples of such 'speciation genes', detected by other approaches¹⁰, are in much older species pairs; they may have evolved after the origin of reproductive isolation, rather than initiating it. So the genes identified in the M-S comparison will be especially interesting. They are likely to contribute to nonrandom mating, the major known barrier to gene flow, and possibly to differential environmental adaptation.

Even though this was a whole-genome survey, it would be premature to assume that all speciation genes fall within the identified islands. The small region detected on chromosome arm 2R (Fig. 1) was only marginally statistically significant. This is, at least partly, because it is small. The methods used would not be able to identify a single speciation gene in an area of high genetic recombination, such as *OdsH* in *Drosophila*⁴, because its signal in terms of the number of differentiated probes would be too weak to detect.

It may be significant that the two major differentiated regions are close to centromeres (Fig. 1), structures by which chromosomes are attached to the spindle during cell division. Centromeres are usually surrounded by areas of reduced recombination, and there is growing evidence that reduced recombination facilitates speciation by preventing the break-up of interacting combinations of variants at specific gene loci¹¹. Restricted recombination also aids the detection of differentiated regions. One of the exciting features of the mosquito system is that patterns such as this can be tested at different stages of speciation. There are seven recognized species in the Gambiae complex that are capable of breeding with each other to a greater or lesser degree, and the microarray approach can be extended to other pairwise comparisons within the group.

The methodology can also be applied to other such pairs of populations in species for which microarrays are available. These systems are rare at the moment, but they will rapidly become more numerous and will provide many insights into the origin of species. ■

Roger Butlin is in the Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK.

e-mail: r.k.butlin@sheffield.ac.uk

Cally Roper is in the Department of Infectious

and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

e-mail: cally.roper@lshtm.ac.uk

1. Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. *PLoS Biol.* **3**, e285 (2005).
2. Wu, C.-I. *J. Evol. Biol.* **14**, 851–865 (2001).
3. Machado, C. A. & Hey, J. *Proc. R. Soc. Lond. B* **270**, 1193–1202 (2003).

4. Ting, C.-T., Tsaur, S.-C. & Wu, C.-I. *Proc. Natl Acad. Sci. USA* **97**, 5313–5316 (2000).
5. Wilding, C. S., Butlin, R. K. & Grahame, J. W. *J. Evol. Biol.* **14**, 611–619 (2001).
6. Borevitz, J. et al. *Genome Res.* **13**, 513–523 (2003).
7. Tripet, F., Touré, Y. T., Dolo, G. & Lanzaro, G. C. *Am. J. Trop. Med. Hyg.* **68**, 1–5 (2003).
8. Gentile, G. et al. *Insect Mol. Biol.* **10**, 25–32 (2001).
9. Lehmann, T. et al. *J. Hered.* **94**, 133–147 (2003).
10. Noor, M. A. F. *Nature* **423**, 699–700 (2003).
11. Butlin, R. K. *Mol. Ecol.* **14**, 2621–2635 (2005).

SYNTHETIC CHEMISTRY

Glycosylation with a twist

Sabine L. Flitsch

Nature has a whole battery of dedicated enzymes to make the complex links between sugar rings — how can synthetic chemists compete? An ingenious approach fills a big gap in the synthetic tool-kit.

Sugars can be linked together in a number of different ways to create chains. The resulting glycoside molecules have numerous biological functions, both by themselves and particularly when attached to proteins and lipids. The chemical reactions involved in making these molecules are formidable, and synthetic chemists have yet to find a set of robust chemical reactions that can generate the full complement of glycosidic linkages. In the *Journal of the American Chemical Society*, Boons and colleagues¹ describe a cunning method that could fill a major gap in the chemical tool-kit required for the efficient synthesis of many glycosides that occur in nature.

In glycosides, each sugar ring is linked through an oxygen atom to its neighbouring ring. The precise configuration of this glycosidic bond determines the molecule's physicochemical properties and biological function. The glycosidic bond occurs in two forms: the attached group can either be below (an α -bond) or above (a β -bond) the plane of the ring (Fig. 1). The chemical similarity of α - and β -bonds presents a problem for the synthesis of glycosides — nature solves this by using a dedicated enzyme (glycosyltransferase) to synthesize each different glycosidic bond. So it seems unlikely that one synthetic glycosylation method will be able to generate all glycosidic linkages. Some very robust reaction conditions have, however, been developed that can be used in different contexts to reliably produce broad classes of glycosides. The most successful of these uses a 'neighbouring group' method.

The classic version of this method uses a protective ester group attached to the carbon-2 position in the sugar ring to direct the formation of the new glycosidic bond to the neighbouring carbon-1 (Fig. 2a, overleaf). During the reaction the oxygen atom of the ester temporarily forms a bond to carbon-1 in the intermediate, termed an 'acetoxonium ion'.

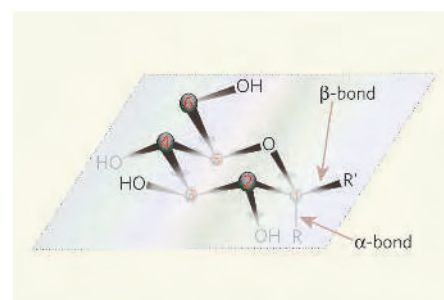


Figure 1 | Configurations of glycosidic bonds using glucose as an example. Carbon atoms are numbered around the ring as shown. R and R' represent different chemical constituents, such as another sugar ring or a hydrogen atom. The bond from carbon-1 to R is an α -bond as it forms below the plane of the ring; the bond from carbon-1 to R' is a β -bond as it forms above the plane of the ring. The R group is *cis* to the nearest –OH group as they are both on the same side of the ring (below the plane); the R' group is *trans* to the nearest –OH group as they are on opposite sides of the ring.

The neighbouring group blocks the lower face of the sugar ring so the glycosidic bond must form from above, making a β -bond. This method has been very useful for making glycosides in which the chemical groups attached to carbon-1 and carbon-2 are on opposite faces of the ring (that is, 1,2-*trans* glycosides). Recently, 1,2-*trans* glycosides have been synthesized by 'one-pot'^{2,3} and even automated routes⁴.

No such robust method is available to synthesize the many glycosides that contain 1,2-*cis* linkages (where the chemical groups attached to carbon-1 and carbon-2 are on the same face of the ring, such as the α -glucosides, α -galactosides or β -mannosides). To address this problem, Boons and colleagues¹ designed a different neighbouring group to create the opposite configuration to that of the classic method (Fig. 2b).



50 YEARS AGO

The same issue of *Smokeless Air* also contains a racy contribution from the pen of a Frenchman, Roger Valvert, entitled "A Frenchman looks at Britain's smoke". A succession of pert reflexions on the Englishman's robustness and toughness in enduring his apparently barbaric forms of heating almost commends the scepticism with which the Beaver Committee report on "Air Pollution" was received last year. We read... "Coal is scarce. One must not waste it. If then this method of heating does not heat, it teaches the English to live in what Continentals would consider cold". Another pleasant gibe runs, "The Englishman to our mind is badly fed and heated, but he has his own flat, or more often house, redecorated every two years, and a well-kept garden. He has a bathroom even if it is arctic. To teach him efficient heating is, therefore, to risk a catastrophe". The author fires his parting shot by suggesting that it would not be a bad thing to observe how the truly sporting spirit helps, that typical characteristic of the Englishman.

From *Nature* 10 September 1955.

100 YEARS AGO

In the months of April and May of this year Vesuvius began to show an increased activity, and in the crater, which was about 80 metres in depth, a small cone began to form; it increased rapidly, and by the middle of May had risen to a height of about 15 metres above the level of the enclosing crater. From May 25 to May 27 violent explosions occurred, which were heard in all the villages on the mountain-side, and were accompanied by the ejection of much red-hot and liquid matter. These explosions ceased almost suddenly on the evening of May 27, and at about 6.45 a small lateral outlet...burst through the north-west flank of the great cone at a height of about 1245 metres, and at the point where a seam in the mountain-side showed where the traces of the last eruption of August 26, 1903, still lingered.

From *Nature* 7 September 1905.

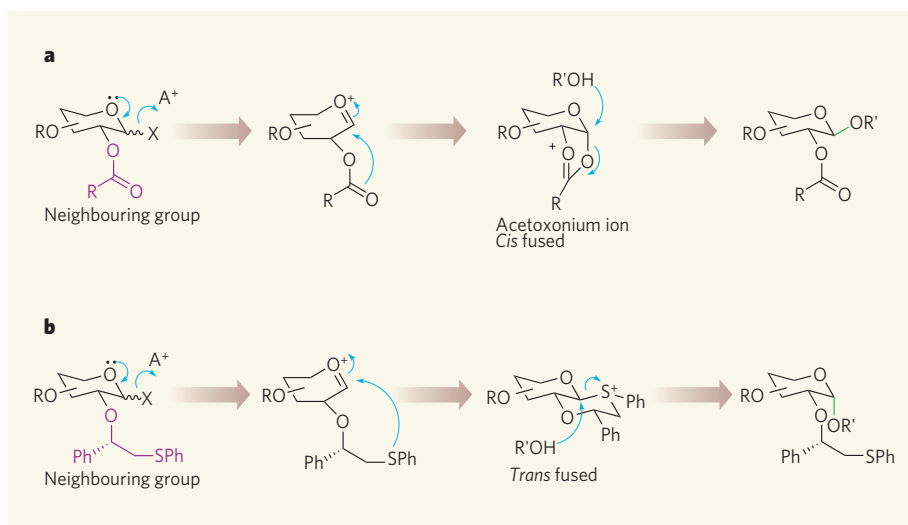


Figure 2 | Sugar twist. **a**, The classic neighbouring group strategy produces a β -glycosidic bond. **b**, Boons and colleagues¹ developed the strategy to generate glycosides with an α -glycosidic bond. R denotes any chemical group; R' is another sugar ring; Ph is a phenol, a six-carbon ring with an -OH group attached; S is a sulphur atom; A is an acid; X is a leaving group; blue arrows show the movement of electrons.

The key difference in their method is that the two-ringed intermediate that forms transiently during the reaction consists of two six-membered rings, rather than one six- and one five-membered ring in the acetoxonium ion. During the reaction shown in Figure 2, the bicyclic ring systems can each be fused in two ways: forming either a *cis*-fused system (with fusion at the same face of both rings; Fig. 2a) or a *trans*-fused system (with fusion at opposite faces; Fig. 2b). Such bicyclic systems have been well studied, and the fundamental chemical principles of their configuration would predict that, although for a six-five ring system *cis* fusion is more energetically favourable, for a six-six ring system, *trans* fusion with one ring twisting around the other is preferred. This switch from *cis* to *trans* changes the outcome of the reaction, resulting in the 1,2-*cis* glycoside.

Boons and colleagues found that using the neighbouring group shown in Figure 2b produced good yields with excellent selectivity for the product with the α -bond. The phenyl component in the neighbouring group improved selectivity. Moreover, it seems that the neighbouring group influences the outcome of the reaction as predicted, because they observed the bicyclic intermediate in the reaction mixture using nuclear magnetic resonance spectroscopy.

A crucial practical consideration for the wider application of this method will be the efficiency with which the new neighbouring group can be installed and cleaved. Very elegantly, the thiophenyl neighbouring group that Boons and colleagues use also facilitates both attachment and cleavage under mild reaction conditions. A slight drawback at the moment, perhaps, is that the protection group needs to be prepared in a rather tortuous four-step procedure.

Boons and colleagues also describe the first application of the method, for the synthesis of a trisaccharide (Gal α 1-3Gal β 1-4GlcNAc) that has been identified as a factor that can trigger acute rejection in transplantation between species. The trisaccharide contains both a 1,2-*cis* and a 1,2-*trans* linkage, and so its synthesis tests the compatibility of both the methods described here — because ultimately one would like to use them in combination. The authors did indeed manage to synthesize the trisaccharide from three suitably protected monosaccharide building blocks in 'one pot' with very high selectivity (no other isomer detected) for the desired glycosidic configurations.

The protocol presented here will need to be tested by many more laboratories on a wide range of targets to establish its robustness and comparison with alternative methods, such as that of intramolecular aglycon delivery⁵. Some gaps remain, such as the attachment of sialic acids, major cell-surface sugars that do not contain groups that are functionally suitable for the present approach. Nevertheless, the glycosylation method described by Boons and colleagues is a big step to greater access to synthetic glycosides.

Sabine L. Flitsch is in the School of Chemistry and the Manchester Interdisciplinary Biocentre, University of Manchester, Oxford Road, Manchester M13 9PL, UK.
e-mail: sabine.flitsch@manchester.ac.uk

- Kim, J.-H., Yang, H., Park, J. & Boons, G.-J. *J. Am. Chem. Soc.* doi:10.1021/ja052548h (2005).
- Douglas, N. L. et al. *J. Chem. Soc. Perkin Trans. 1* 51-66 (1998).
- Koeller, K. M. & Wong, C.-H. *Chem. Rev.* **100**, 4465-4493 (2000).
- Seeberger, P. H. *Chem. Commun.* 1115-1121 (2003).
- Fairbanks, A. J. *Synlett* 1945-1958 (2003).

MOLECULAR PHYSIOLOGY

Intimate contact enables transport

Baruch I. Kanner

Sodium-coupled neurotransmitter transporters are essential for neurons to communicate. The high-resolution crystal structure of a bacterial relative hints at how this family of transporters works.

Ion-coupled transporters are essential for all forms of life. These molecular machines carry their cargoes across membranes to help accumulate foodstuffs in cells, maintain cellular pH levels and facilitate communication between nerve cells. Transporters must move their substrates (sugars, amino acids, neurotransmitters, and so on) against sometimes huge concentration gradients. This is accomplished using the energy stored in the electrochemical ion gradients that are maintained across cell membranes. On page 215 of this issue Yamashita *et al.*¹ report the crystal structure of a bacterial leucine transporter, the first structure of a member of the large Na^+/Cl^- -dependent neurotransmitter transporter family. It shows that one of the two sodium ions bound in the transporter's binding pocket comes into direct contact with the leucine molecule being carried across the membrane.

Transporters generally function by exposing their binding sites alternately to either side of the membrane — catching up their cargo on one side and releasing it on the other. A widely accepted theory proposes that this can be accomplished using two gates, with only one open at a time, just like the locks in a waterway². Support for this idea comes from crystal structures of transporters, which invariably show a cavity in the transporter closed off from the aqueous space on either or both sides of the membrane.

The binding pocket of ion-coupled transporters also has binding sites for the ion that powers the transport process. So the ion and the substrate are transported together — although sometimes in opposite directions — such that the energy released as the ion moves down its gradient is used to power the uphill movement of the substrate. A question yet to be resolved is how the 'driving' ion and the 'driven' substrate move through an ion-coupled transporter.

Yamashita *et al.*¹ studied LeuT_{Aa}, a leucine transporter from the bacterium *Aquifex aeolicus*, which is a member of the family of Na^+/Cl^- -dependent neurotransmitter transporters^{3,4}. Some of the best-known family members function in the central nervous system, where they carry neurotransmitters, the brain chemicals used to signal across the synaptic junctions between neurons. The transporters clear neurotransmitters from the synaptic cleft, terminating the signal and priming the neurons for the next one. Consequently, chemicals that inhibit these proteins — such as cocaine and Prozac —

have profound effects on brain function.

Crystal structures of various transporters from other families^{5–8} hint that nature has found several solutions to alternating the access of the binding pocket to each side of the membrane. Now Yamashita *et al.*¹ have resolved the structure of LeuT_{Aa} at exceptionally high resolution for a membrane protein (1.65 Å). It reveals not only a completely new protein fold, but also a crystal-clear view of the binding pocket — including the driven substrate and the two driving sodium ions (Fig. 1). The sodium ions in the binding pocket are both close to the leucine, with one of them being in direct contact through the carboxyl group (Fig. 1, inset). So it seems that, at least in this transporter and presumably in all the amino-acid-transporting members of this protein family, the coupling between the driving ion and the driven substrate is as direct as it could be.

This direct coupling is an ingenious solution to minimizing leaks where the ion and/or the

substrate might permeate through the transporter independently. This mechanism has been proposed previously on the basis of indirect evidence from transporters of other families^{9,10}, and it may turn out to be used by many transporters. Nevertheless, direct contact may not occur in all transporters; for instance, not all the substrates of transporters related to LeuT_{Aa} have carboxyl groups. In these cases it seems that the carboxyl group is provided by a unique aspartate residue located on transmembrane domain 1 (TM1).

The structure of the functional unit, the LeuT_{Aa} monomer, shows several features known from other transporter structures. One is an internal structural repeat in the LeuT_{Aa} monomer, such that TM1–TM5 and TM6–TM10 can be superimposed on each other by rotation around a pseudo twofold axis located in the plane of the membrane. The interface of these repeats forms the binding pocket of the transporter (Fig. 1).

Another feature is the unwinding of

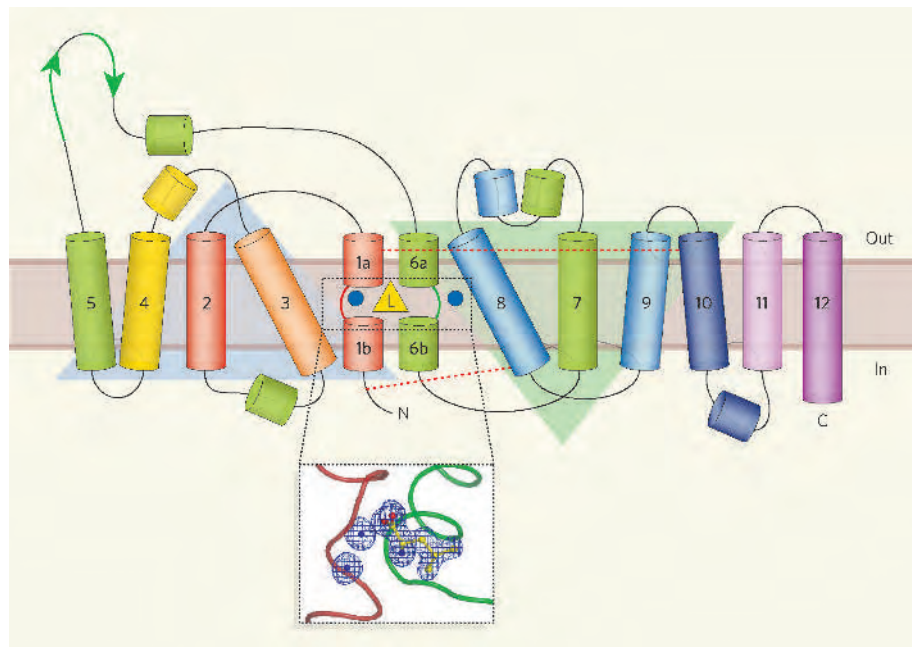


Figure 1 | The LeuT_{Aa} topology revealed by Yamashita and colleagues¹. Transmembrane domains (TMs) are numbered 1–12, and the oppositely oriented structural repeats encompassing TM1–TM5 and TM6–TM10 are shown as blue and green triangles. TM1 and TM6 are unwound halfway through the membrane, to form the binding pocket for the sodium ions (blue circles) and the leucine cargo (yellow triangle). This structure has its binding sites occluded. The two dashed red lines connect the approximate positions of amino acids that interact as ion pairs to form parts of the external and internal gates. The residues involved are close together in three dimensions. Presumably they reciprocally dissociate and associate such that the binding pocket is accessible to the extracellular side when the extracellular pair is dissociated and the intracellular pair is associated, and vice versa. Inset, the binding pocket showing the actual electron densities representing leucine (carbon, yellow; oxygen, red; nitrogen, blue) and the two sodium ions. (Figure adapted from ref. 1.)

membrane-spanning domains, which was first observed in the calcium pump¹¹. In LeuT_{AA}, TM1 and TM6 are antiparallel to each other, and have breaks in their helical structure approximately halfway across the membrane (Fig. 1). These breaks expose main-chain carbonyl oxygen and nitrogen atoms for hydrogen bonding and ion binding. Residues on TM3, TM7 and TM8 also contribute to the binding of sodium and leucine. Some of these residues had already been implicated in ion and/or substrate binding by functional studies of mutants of several of the neurotransmitter transporters (cited in ref. 1). Therefore, it appears that the structure reported by Yamashita *et al.* is a physiologically relevant conformation of the transporter.

In this structure, the binding pocket is occluded — the external and internal gates are closed. Two ion pairs, one between the extracellular ends of TM1 and TM10 and the other between the intracellular ends of TM1 and TM8, contribute to these gates (Fig. 1). The crystal of LeuT_{AA} also contains a chloride ion (not shown on the figure), but this is not located in the binding pocket. Indeed, in LeuT_{AA}, leucine transport is dependent on sodium but not on chloride¹. In contrast, in GAT-1, another member of the family, the neurotransmitter GABA is transported together with sodium and chloride ions¹². So although the overall structure of the neurotransmitter transporters is expected to be similar to that of LeuT_{AA}, there will be variations.

In the future, the LeuT_{AA} structure will be useful in designing drugs that specifically inhibit the neurotransmitter transporters. Obtaining more 'snapshot' structures representing different transporter conformations will shed light on the most fundamental question: which conformational changes occur during the transport cycle? Or, in terms of the 'lock' model, how is the opening of the external gate coupled to the closing of the internal one, and vice versa?

Baruch I. Kanner is in the Department of Biochemistry, Hebrew University Hadassah Medical School, PO Box 12272, Jerusalem 91120, Israel.
e-mail: kannerb@cc.huji.ac.il

1. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. *Nature* **437**, 215–223 (2005).
2. Jardetzky, O. *Nature* **211**, 969–970 (1966).
3. Guastella, J. *et al. Science* **249**, 1303–1306 (1990).
4. Pacholczyk, T., Blakely, R. D. & Amara, S. G. *Nature* **350**, 350–354 (1991).
5. Abramson, J. *et al. Science* **301**, 610–615 (2003).
6. Huang, Y., Lemeroux, M. J., Song, J., Auer, M. & Wang, D.-N. *Science* **301**, 616–620 (2003).
7. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. *Nature* **431**, 811–818 (2004).
8. Hunte, C. *et al. Nature* **435**, 1197–1202 (2005).
9. Hama, H. & Wilson, T. H. *J. Biol. Chem.* **268**, 10060–10065 (1993).
10. Zhang, Y. & Kanner, B. I. *Proc. Natl Acad. Sci. USA* **96**, 1710–1715 (1999).
11. Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. *Nature* **405**, 647–650 (2000).
12. Keynan, S. & Kanner, B. I. *Biochemistry* **27**, 12–17 (1988).

ENVIRONMENTAL SCIENCE

Carbon unlocked from soils

E. Detlef Schulze and Annette Freibauer

Changes in climate and land use are implicated as the main factors in the large-scale loss of carbon from soils in England and Wales over the past 25 years. The same picture is likely to apply much more broadly.

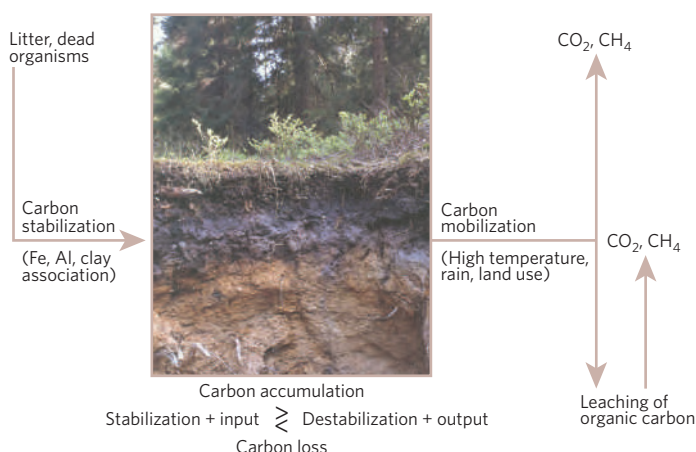
Soils are major players in the carbon cycle — globally, they store the equivalent of about 300 times the amount of carbon now released annually through the burning of fossil fuels. It is generally assumed that most of the carbon locked up in soils is inert, and stays there. But as Bellamy *et al.* report on page 245 of this issue¹, soil carbon may be more vulnerable to changing climate and patterns of land use than expected.

The carbon involved is known as soil organic carbon (SOC; Box 1). Bellamy *et al.* describe how they have determined the changes in SOC stocks in the top 15 cm of soils in England and Wales during the past 25 years. Their estimates are based on a soil inventory of almost 6,000 sites across all types of land use, resampled on a systematic grid from a

pre-existing larger inventory. They find SOC losses of an alarming magnitude. Extrapolating to the entire United Kingdom, Bellamy *et al.* estimate annual losses of 13 million tonnes of carbon. This is equivalent to 8% of the UK emissions of carbon dioxide in 1990, and is as much as the entire UK reduction in CO₂ emissions achieved between 1990 and 2002 (12.7 million tonnes of carbon per year).

These losses thus completely offset the past technological achievements in reducing CO₂ emissions, putting the United Kingdom's success in reducing greenhouse-gas emissions in a different light. Under the Kyoto Protocol, however, countries are not obliged to account for changes in the stock of soil carbon. So an effective climate policy will require a more comprehensive approach that includes all

Box 1 Soil carbon in context



Organic carbon is stored in the top layers of mineral soil as humus or above the mineral soil as peat or litter. This organic material is by no means in equilibrium — neither the carbon concentrations nor the depth of the soil layers are constant, although changes generally occur very slowly. Soils receive dead organic material, known as litter, mainly from the plant cover. This material is decomposed by the soil biota and partly mineralized, and is subsequently released to the atmosphere in the form of carbon dioxide and methane, or by leaching into groundwater¹³.

The subtle balance between input and output determines whether a soil is accumulating or losing carbon. Soil organic carbon (SOC) consists of diverse compounds with different chemical and physical properties; for scientific purposes, SOC is divided into an active and a passive pool, the latter being more resilient to further

degradation and possibly existing in soil for hundreds to thousands of years. All factors that reduce biological activity, and that stabilize SOC by physical protection or binding to clay silicates or metals, will promote accumulation; factors that increase biological activity and destabilization encourage degradation.

The interplay of these factors is highly complicated. For example, in humid regions, given an adequate supply of moisture, global warming may increase microbial activity and accelerate SOC mobilization; in drier areas, the converse may apply. Changing patterns of land use can also have significant effects. Losses of SOC occur when natural ecosystems are cultivated — because of degradation of soil fertility, intensified soil disturbance and reduced carbon input^{10,11}. If conservation measures are applied to degraded soils, however, the SOC content can be maintained or enhanced¹⁴.

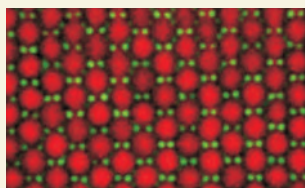
E.D.S. & A.F.

PHYSICAL CHEMISTRY

Isn't it ionic

Microscopic particles dispersed in solvents — known as colloidal suspensions — are attractive models for studying a wide range of phase transitions and nucleation phenomena. The suspended particles can be observed directly in three-dimensional space, and the interactions between them are easily modified. However, processes involving that seemingly most simple and ubiquitous of inorganic solids — ionic crystals formed from oppositely charged atoms — cannot be mimicked by colloids. Charged colloidal particles don't 'do ionic'; they prefer to aggregate instead.

But Mirjam E. Leunissen and colleagues, writing in this issue (*Nature* **437**, 235–240; 2005), show that colloids can be coaxed into forming ionic crystals after all. The authors observed that if salt is added to polymer spheres dispersed in an organic solvent mixture, the charge on the spheres can be controlled and the electrostatic interactions between them can be regulated. This enabled the preparation of binary mixtures of colloids that carried small, opposite charges and readily formed ionic crystals. When an electric field was applied, the crystal melted, and particles of



opposite charge moved towards opposite electrodes.

The charged colloidal particles therefore clearly resemble ionic species. But there are differences. In particular, a diffuse layer of 'counter-ions' surrounds each particle, forming an overall charge-neutral unit that participates in the growth of the crystal. So the structure of the colloidal crystals is not dictated by charge neutrality, as in atomic systems, leaving the authors free to create remarkable

new binary structures. One example, a crystal comprising particles of positive (green, radius 0.36 μm) and negative (red, radius 1.16 μm) charge in the ratio 6:1, is shown in the image.

Colloidal crystals can also form from charged spheres made of different materials, such as a polymer and silica. It is then straightforward to burn the polymer spheres away to give all-silica structures. Given the ease with which these structures grow into large crystals, ionic colloids should prove an alluring proposition for those creating advanced materials such as photonic crystals.

Magdalena Helmer

major carbon sources and sinks in the biosphere², with emphasis on protecting existing pools of stored carbon.

Bellamy and colleagues' observations are remarkable, and fall into four categories. First, consistent losses of SOC occurred independently of soil properties, challenging our knowledge about SOC stability. A new hypothesis suggests that the stability of SOC depends on the diversity and activity of soil microorganisms — if microbes have adequate energy resources, they can break down any organic structure irrespective of its physico-chemical stabilization³.

Second, the losses were proportional to carbon concentration, which implies a first-order decay of a homogeneous pool. This contradicts the view that SOC in carbon-rich soils contains a higher fraction of stable carbon than does that in carbon-poor soils. But it is unclear from which SOC fraction the lost carbon originated.

Third, SOC losses occurred in soils under all land-use conditions, and there is no apparent single factor other than climate change that could degrade non-agricultural soils. Fourth, the observation is contrary to the view⁴ that soils are an ultimate, large-scale sink for carbon.

The overall net loss of carbon identified by Bellamy *et al.* is unequivocal. But the authors can only speculate about where the carbon has gone. The reported changes in the top 15 cm correspond to average losses of 125 g C m⁻² yr⁻¹, ranging from 66 g C m⁻² yr⁻¹ uptake in extremely carbon-poor soils to 550 g C m⁻² yr⁻¹ loss in carbon-rich (peat) soils. The SOC content may also have changed in soil layers not studied by Bellamy and colleagues. This may affect the magnitude of the overall carbon loss but is unlikely to negate the findings. Leaching of dissolved organic carbon cannot explain more than 10% of the loss^{5,6}, and it is likely that most of the rest has been converted to CO₂.

Direct measurements of carbon flux do not yet provide a complete picture of the carbon cycle in ecosystems, and the 'mass balance' approach of Bellamy *et al.* is the only available evidence of change in the SOC pool. But the reasons for the loss remain unclear, irrespective of the pathways by which carbon is being lost. Re-inspecting the results, we think that the land-use factor has played a role — for example, only alteration in land use and gradual changes in land management can explain why croplands lost more carbon than other areas. Major land-use changes, such as the afforestation of carbon-rich soils, are not highlighted separately in Bellamy and colleagues' study.

Climate variation seems to be the second factor. Modelling studies suggest that significant changes in SOC stem from variations in precipitation and temperature on timescales of decades^{7,8}. But these conclusions are based only on laboratory studies and small-scale experiments. According to our current understanding of the sensitivity of soil respiration to warming, increased temperature alone seems to be too weak a driver. Carbon-rich soils lost most SOC. In the United Kingdom, these soils are often very wet and can easily be affected by changing precipitation patterns. The relative effects and combined impact of increased atmospheric CO₂, warming, nitrogen deposition and altered precipitation are still disputed.

Bellamy and colleagues' observations are sure to stimulate further investigations to clarify the main uncertainties. International research projects such as CarboEurope, the Global Carbon Project and others⁹ are already under way with the aim of providing a better understanding of carbon balance in the biosphere. But this work is by no means straightforward. Soils are a nightmare to work with — they are dirty and highly varied in composition, and if a sample has been taken for analysis, that spot cannot be resampled.

The scientific and political implications

of the new findings¹ are considerable. The process of carbon loss from soils has been most comprehensively documented in the United Kingdom, both at regional level and under all forms of land use. But there have also been repeated warning signals from soil surveys in China¹⁰, Finland¹¹ and Flanders¹². These, however, attribute most of the SOC loss to changes in land use and management. In contrast, Bellamy *et al.* provide the first hint that regional climate variation may be contributing to a surprisingly large release of CO₂ from soils to the atmosphere. Further research into the carbon cycle and on reducing CO₂ emissions must take full account of areas where large pools of organic carbon are stored — or are being released. If we intend to stabilize the climate, such areas require much more serious consideration. ■

E. Detlef Schulze and Annette Freibauer are at the Max-Planck Institute for Biogeochemistry, PO Box 1001 64, 07701 Jena, Germany.
e-mail: Detlef.Schulze@bgc-jena.mpg.de

1. Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M. & Kirk, G. J. D. *Nature* **437**, 245–248 (2005).
2. German Advisory Council on Global Change. *World in Transition: Towards Sustainable Energy Systems* (Earthscan, London, 2004).
3. Gleixner, G. *et al.* in *Global Biogeochemical Cycles in the Climate System* (eds Schulze, E. D. *et al.*) 201–216 (Academic, San Diego, 2001).
4. Schulze, E.-D. (ed.) *Carbon and Nitrogen Cycling in European Forest Ecosystems* (Springer, Berlin, 2000).
5. Siemens, J. *Science* **302**, 1681 (2003).
6. Worrall, F., Reed, M., Warburton, J. & Burt, T. *Sci. Total Environ.* **312**, 133–146 (2003).
7. Tian, H. *et al.* *Tellus B* **51**, 414–452 (1999).
8. Betts, R. A. *et al.* *Theor. Appl. Climatol.* **78**, 157 (2004).
9. www.carboeurope.org; www.globalcarbonproject.org; http://daac.ornl.gov/FLUXNET; www.isse.ucar.edu/nacp/
10. Wang, S. *et al.* *Tellus B* **55**, 416–427 (2003).
11. Finnish Ministry of the Environment. *Finland's Third National Communication under the United Nations Framework Convention on Climate Change* (Environ. Ministry, Helsinki, 2001).
12. Sleutel, S. *et al.* *Glob. Change Biol.* **9**, 1193–1203 (2003).
13. Schulze, E. D., Beck, E. & Müller-Hohenstein, K. *Plant Ecology* (Springer, Heidelberg, 2005).
14. Follett, R. F. *Soil Tillage Res.* **61**, 77–92 (2001).

BRIEF COMMUNICATIONS

No hormonal response in tied fights

Fish androgens may start to surge only when there seems to be a good chance of winning a contest.

Androgens are the principal sex steroids controlling reproduction and aggression in male fish¹, but their production can also be affected by social interactions^{2,3}. Here we show that androgen concentrations are not significantly increased in cichlid fish (*Oreochromis mossambicus*) that are fighting their own image in a mirror, despite their aggressive behaviour towards the virtual intruder. Our results indicate that the hormonal response normally triggered in male contests is not induced under these circumstances by the act of fighting itself, and that it may therefore depend on some indicator of relative fighting ability that cannot be delivered by a mirror-image challenger.

Fish do not recognize their own image in a mirror and so attack it as though it were an intruder⁴. However, there are no pointers (such as differences in rank, strength, injury or fatigue) that can be used to assess the outcome of the attack (winning versus losing) for a fish in this situation. To test whether information on interaction outcome is necessary to trigger an androgen response in fighting fish, we compared androgen concentrations in the urine of fish after mirror-mediated challenge with those in urine from non-fighting controls.

Male cichlids of comparable size and age, raised in similar conditions, were kept in social



Duel action: a cichlid attacks its reflection.

isolation for seven days before the experiment to minimize inter-individual variation in behavioural and androgen responses that arise from differences in social status⁵. On the day of the experiment, the fish were divided into two groups: one was presented with a mirror at one end of the tank (experimental group; $n = 17$); the other had a sheet of transparent glass instead to control for the presence of a novel object in the aquarium (control group; $n = 14$). We recorded the fighting behaviour of the fish (aggressive displays and attacks) for 20 min after the first interaction with the mirror or glass by using a focal continuous recording method⁶. The principal fish androgens¹, testosterone and 11-keto-testosterone, were measured by specific radioimmunoassay of urine

collected 60 min before the test (baseline value) and 30 min, 2 h and 6 h after the end of the test.

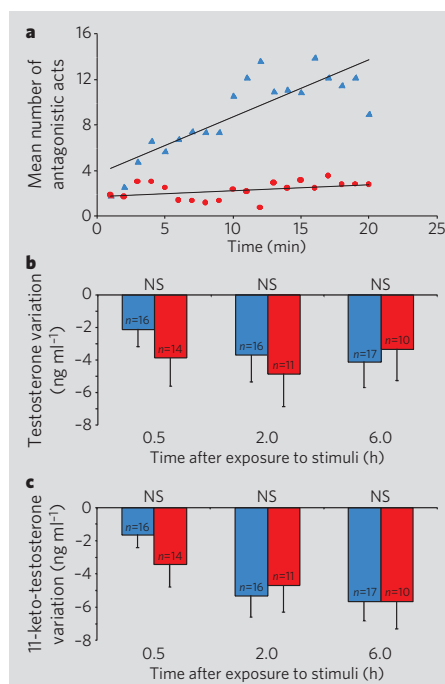
Individuals reacted very aggressively to their own image in the mirror, whereas the presence of non-reflective glass in the tank did not provoke a combative response in fish from the control group. During the course of the attacks, the aggression of subjects in the experimental group escalated towards their mirror images as they apparently tried to resolve the 'contest' (Fig. 1a).

However, despite the marked degree of aggressive behaviour, no significant increase in urine androgens was detected in the experimental relative to the control group (Fig. 1b, c). This contrasts with the increased production of androgens in *O. mossambicus* males in response to territorial intrusions by conspecific males⁷, and with the androgen increase in male spectators that are not even directly participating in a fight⁸. Although there is a reduction in androgen synthesis in fish that lose fights and become subordinates⁵, this cannot account for the drop in androgens in the experimental group, which is similar to that in the non-fighting control group and is in the range expected for the time of day (morning) at which the measurements were made.

We conclude that information on the likely outcome of a contest is required before the androgen response is triggered in a combatant, and that escalation of the fight occurs independently of circulating androgen levels. This may be an adaptive mechanism that allows individuals to mount an androgen response for the purpose of controlling their social status after they have assessed the relative fighting ability of their challenger.

Rui F. Oliveira*, Luis A. Carneiro*, Adelino V. M. Canário†

*Unidade de Investigação em Eco-Etologia, Instituto Superior de Psicologia Aplicada,



1149-041 Lisboa, Portugal

e-mail: ruioi@ispa.pt

†CCMar, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

1. Borg, B. *Comp. Biochem. Physiol. C* **109**, 219–245 (1994).
2. Oliveira, R. F. *Adv. Study Behav.* **34**, 165–239 (2004).
3. Wingfield, J. C., Hegner, R. E., Dufty, A. M. & Ball, G. F. *Am. Nat.* **136**, 829–846 (1990).
4. Rowland, W. J. *Environ. Biol. Fish.* **56**, 285–305 (1999).
5. Oliveira, R. F., Almada, V. C. & Canário, A. V. M. *Horm. Behav.* **30**, 2–12 (1996).

6. Martin, P. & Bateson, P. *Measuring Behaviour: An Introductory Guide* (Cambridge Univ. Press, Cambridge, 1993).
7. Hirschenhauser, K., Taborsky, M., Oliveira, T., Canário, A. V. M. & Oliveira, R. F. *Anim. Behav.* **68**, 741–750 (2004).
8. Oliveira, R. F., Lopes, M., Carneiro, L. A. & Canário, A. V. M. *Nature* **409**, 475 (2001).
9. Scott, A. P. & Sumpter, J. P. *Gen. Comp. Endocrinol.* **73**, 46–58 (1988).
10. Scott, A. P., Mackenzie, D. S. & Stacey, N. E. *Gen. Comp. Endocrinol.* **56**, 349–359 (1984).

Competing financial interests: declared none
doi:10.1038/437207a

DYNAMICAL PHENOMENA

Walking and orbiting droplets

Small drops can bounce indefinitely on a bath of the same liquid if the container is oscillated vertically at a sufficiently high acceleration¹. Here we show that bouncing droplets can be made to 'walk' at constant horizontal velocity on the liquid surface by increasing this acceleration. This transition yields a new type of localized state^{2–5} with particle–wave duality: surface capillary waves emanate from a bouncing drop, which self-propels by interaction with its own wave and becomes a walker. When two walkers come close, they interact through their waves and this 'collision' may cause the two walkers to orbit around each other^{6–8}.

The bouncer transition to walking is continuous and occurs when the vertical acceleration of the bath, γ_m , reaches a critical threshold, γ_m^c . Below γ_m^c , the drops bounce with no horizontal motion. Above γ_m^c , bouncing drops acquire a rectilinear motion along the surface of the bath (Fig. 1a–c). Their velocity V_w is constant (0–20 mm s^{−1}) and increases with γ_m .

Why do the drops start walking? This phenomenon occurs below, but near, the onset of the Faraday instability, a point at which the surface becomes spontaneously wavy. In this regime, the vertical motion of a drop becomes subharmonic, with a period that is double that of the forcing. As a result, it emits a damped Faraday wave. The drop undergoes successive identical parabolic jumps that are locked with its wave. Each jump brings the drop into collision with the side of the central bulge of the wave generated by the previous collision (Fig. 1a). This collision with an inclined surface generates a non-zero horizontal impulse, which can be translated as an equation for the drop's horizontal motion, averaged over a period π/ω_0 of the subharmonic vertical motion

$$m \frac{d^2x}{dt^2} = a \sin\{(\pi k/\omega_0) dx/dt\} - b dx/dt \quad (1)$$

where m is the drop's mass, a is about 10^{-6} N, k is the wavenumber, and b is about 10^{-6} N m^{−1} s. The left-hand side of equation (1) represents the inertia of the drop; the first term on the right-hand side accounts for the effective force due to the inclined surface, and the second for viscous damping during the collision. Equation

(1) predicts the observed continuous transition of the droplet from stationary to walking when $a > b\omega_0/(\pi k)$.

When walkers coexist in a cell, they inevitably collide. These 'collisions' do not involve any contact between the drops but only a deflection of their horizontal trajectories, when the wave generated by a drop affects the horizontal velocity of the other one. The main parameter characterizing this collision is d_c , the minimal distance of approach of the two drops; depending on the value of d_c , the walkers either attract or repel each other. Attraction leads to a twin-star-like orbiting motion of the drops (Fig. 1d, and see movie in supplementary information). The diameters of the orbits take discrete values d_n^{orb} , which self-adapt to the forcing frequency^{9,10}. The orbital diameters are slightly smaller than an integer multiple of the Faraday wavelength (λ_F), or $d_n^{\text{orb}} = (n - \varepsilon)\lambda_F$ when the drops bounce in phase. They are $d_n^{\text{orb}} = (n + 1/2 - \varepsilon)\lambda_F$ when the drops bounce in antiphase; the offset, $\varepsilon = 0.2 \pm 0.02$, is such that when a drop collides with the surface, it falls on the inward slope of the wave emitted by the other. This provides the centripetal force needed for the orbital motion. For other values of d_c , each drop falls on the outward slope of the wave of the other, which causes a repulsion.

We have shown that walkers can behave as billiard balls, undergo scattering collisions or form circular orbits, and can even display complex three-body motion (results not shown). The variety of these phenomena can be explained by interaction through waves and by generalizing equation (1) to two or more drops (the resulting equations yield the same quantification of orbits and numerical trajectories, which are very similar to the experimental collisions; S. P. *et al.*, manuscript in preparation). In this system, real particles experience the same non-local interaction as nonlinear waves.

Y. Couder*, S. Protière*, E. Fort†, A. Boudaoud‡

*Matières et Systèmes Complexes, and †Matériaux et Phénomènes Quantiques, Université Paris 7 Denis Diderot, 75005 Paris, France

‡Laboratoire de Physique Statistique, 24 rue Lhomond, 75231 Paris, France
e-mail: boudaoud@lps.ens.fr

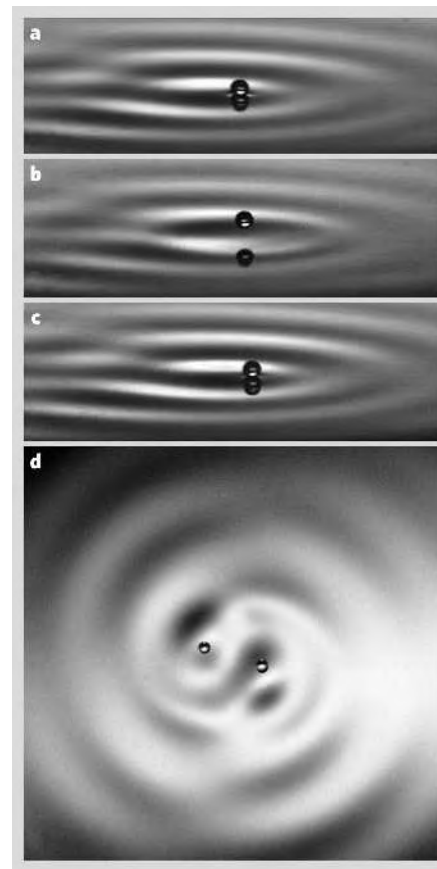


Figure 1 | Behaviour of silicon oil droplets on a bath of silicon oil when it is oscillated vertically.

Experimental parameters: oil viscosity, 20×10^{-3} Pa s; forcing frequency, $\omega_0/2\pi = 80$ Hz; diameter of droplets $D \approx 0.65$ mm; forcing acceleration, $\gamma_m/g \approx 3.9$ (where g is the acceleration due to gravity). **a–c**, Photographs showing the motion of a single drop in interaction with its own localized Faraday wave on the liquid surface. The drop's motion is composed of a series of identical parabolic jumps, each jump bringing the drop into collision with the forward side of the central bulge of the wave generated by the previous collision. **d**, Photograph of two orbiting drops and associated waves. The horizontal motion is in a twin-star-like orbit of diameter $d_n = 5.8$ mm. (For movies, see supplementary information.)

1. Couder, Y., Fort, E., Boudaoud, A. & Gautier, C. H. *Phys. Rev. Lett.* **94**, 177801 (2005).
2. Fauve, S. & Thual, O. *Phys. Rev. Lett.* **64**, 282–284 (1990).
3. Aranson, I. S., Gorshkov, K. A., Lomov, A. S. & Rabinovich, M. I. *Physica D* **43**, 435–453 (1990).
4. Lioubashevski, O., Arbell, H. & Fineberg, J. *Phys. Rev. Lett.* **76**, 3959–3962 (1996).
5. Umbanhowar, P. B., Melo, F. & Swinney, H. L. *Nature* **382**, 793–796 (1996).
6. Stegeman, G. I. & Segev, M. *Science* **286**, 1518–1523 (1999).
7. Moskalenko, A. S., Liehr, A. W. & Purwins, H. G. *Eur. Phys. Lett.* **63**, 361–367 (2003).
8. Liehr, A. W., Moskalenko, A. S., Astrov, Y. A., Bode, M. & Purwins, H. G. *Eur. Phys. J. B* **37**, 199–204 (2004).
9. Boudaoud, A., Couder, Y. & Ben Amar, M. *Eur. Phys. J. B* **9**, 159–165 (1999).
10. Boudaoud, A., Couder, Y. & Ben Amar, M. *Phys. Rev. Lett.* **82**, 3847–3851 (1999).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none
doi:10.1038/437208a

BRIEF COMMUNICATIONS ARISING online
♦ www.nature.com/bca see Nature contents.

MALARIA RISK

Estimating clinical episodes of malaria

Arising from: R. W. Snow, C. A. Guerra, A. M. Noor, H. Y. Myint & S. I. Hay *Nature* **434**, 214–217 (2005)

Estimates of the disease burden caused by malaria are crucial for informing malaria control programmes. Snow and colleagues claim that their estimate of 515 million cases of malaria caused by *Plasmodium falciparum* globally is up to 50% higher than that reported by the World Health Organization (WHO), and 200% higher for areas outside Africa¹. However, this comparison refers to the WHO's estimates from 1990 and 1998, and not to the range of 300 million to 500 million that the WHO has used since 2000 (ref. 2). Both groups agree that the burden of malaria disease outside Africa, especially in South Asia, is greater than was estimated in the 1990s.

A new global map of populations living at risk of malaria transmission, produced in 2004 and funded in part by the WHO's Roll Back Malaria department (WHO/RBM), forms the basis of new estimates both by Snow and colleagues¹ and the WHO/RBM³. Both current sets of estimates reflect the consensus that cases recorded and reported in national health information systems capture far less than the full burden of malaria in most parts of the world.

In October 2004, a group of independent experts reviewed the WHO/RBM's new estimation method and concluded that, even with the best available data, it is preferable to present any malaria incidence estimate as a range⁴. An important reason for this conclusion is that the research studies used as input are normally

conducted in areas where malaria transmission is greatest, and during the season of peak malaria transmission. Extrapolation from these studies may therefore result in a picture that is not truly representative of the entire region. In addition, the population-at-risk map may not be totally accurate, given fluctuations in malaria transmission patterns in response to environmental change, development and vector control⁴.

The WHO's global burden estimate for 2004 of 350 million to 500 million cases, of which 270 million to 400 million are due to infection by *P. falciparum*³, is generally consistent with the 300 million to 660 million range for *P. falciparum* proposed by Snow and colleagues¹. The somewhat smaller range estimated by the WHO is likely to be due to our inclusion of the impact of preventive interventions (insecticide-treated mosquito nets and indoor residual spraying), the coverage of which has increased since 2000. Further, for selected areas with highly unstable malaria transmission, the WHO's estimate takes into account case numbers reported through routine health information systems, which often reflect both passive and active case detection and, in these settings, may provide a more reliable and up-to-date picture than extrapolations made from published research in sometimes distant sites.

The WHO is working with countries to improve their capacity for collecting the data

required for future national-level incidence estimates. Data will be obtained from national surveys and sentinel surveillance sites, as well as from health information systems. These efforts should improve the precision of burden estimates at all levels and allow us to assess, in the second half of this decade, whether the ongoing increase in coverage of effective prevention and treatment measures are bringing us closer to the global goal of reducing the burden of malaria by half by the year 2010.

Bernard L. Nahlen*, **Elaine L. Korenromp***, **John M. Miller***, **Kenji Shibuya†**

World Health Organization, *Roll Back Malaria Department and †Evidence for Information and Policy Department, 1211 Geneva, Switzerland
e-mail: korenrompe@who.int

1. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. *Nature* **434**, 214–217 (2005).
2. WHO Expert Committee on Malaria: 20th Report (World Health Organization, Geneva, 2000).
<http://www.rbm.who.int/docs/ecr20.pdf>
3. Korenromp, E. L. for the RBM MERG Task Force on Malaria Morbidity: *Malaria Incidence Estimates at Country Level for the Year 2004 — Proposed Estimates and Draft Report* (World Health Organization, Roll Back Malaria, Geneva, 2005).
http://mosquito.who.int/docs/incidence_estimations2.pdf
4. *Minutes of the First MERG Task Force Meeting on Malaria Morbidity* (World Health Organization, Roll Back Malaria, Geneva, 2004). http://rbm.who.int/partnership/wg/wg_monitoring/docs/MorbidityTaskForce_meetingOct04_2004.pdf

doi:10.1038/nature04178

Reply: Snow *et al.* reply to this communication (doi:10.1038/nature04180).

MALARIA RISK

Estimation of the malaria burden

Arising from: R. W. Snow, C. A. Guerra, A. M. Noor, H. Y. Myint & S. I. Hay *Nature* **434**, 214–217 (2005)

Accurate estimates of the global burden of malaria are important for planning, monitoring and advocacy. Snow *et al.*¹ attempt to address the shortcomings of previous estimates of the incidence of malaria caused by *Plasmodium falciparum* by combining current and historical data. However, we believe that the design of their model and its inputs have led to a significant overestimate of the malaria burden outside Africa — as in the example of the World Health Organization (WHO) western Pacific region (WPR), for which their model predicts 60 times the 2002 incidence reported by national malaria-control programmes².

The reliance by Snow *et al.* on three broad categories of malaria risk — hypo-, meso-, and hyper/holoendemic malaria — would seem to

be the likely basis for this overestimate. The 1968 map³ that Snow *et al.* used to obtain these endemic zones was developed from estimates of maximum malaria prevalence between the late nineteenth century and the 1960s and included some *P. vivax* infections. Snow *et al.* laid this over a current (2002) WHO map of malaria distribution, modified according to travel advice and some national statistics. However, the scale of this WHO map indicates only broad areas where transmission may be found. The reduction of prevalence in highly endemic areas by one step from its historical maximum to account for control methods and deforestation, and the further reduction of one step in urban areas of more than 1 million inhabitants¹, seems arbitrary and assumes that

transmission in smaller cities and towns is equal to that in forested areas. Hypoendemic areas outside major cities remain unchanged from the pre-1968 map, unless national figures were known to show absence of transmission. Despite the stated methods, holo/hyperendemic zones still also appear in calculations¹.

The predominant vectors of malaria in southeast Asian countries of the WPR are associated with forests⁴. In the past 50 years, deforestation has removed much of this habitat⁵. Furthermore, although the proportion of the population living in urban areas has greatly increased, urban and peri-urban malaria is uncommon in Asian countries, and virtually absent in the Philippines, Malaysia and the Mekong region. Many people currently living in areas that were highly endemic for malaria on the Lysenko and Semashko map³ are therefore at low or no risk of malaria. Malaria incidence in remaining transmission areas is geographically highly heterogeneous^{6–9}.

We conclude that the prediction by Snow *et al.* of 15.03 million cases in the WPR, based predominantly on an estimated population of

63.4 million at risk of 171 *P. falciparum* malaria cases per 1,000 per year ("mesoendemic")¹, is not consistent with the epidemiology of malaria in the region. Apart from Vanuatu, the Solomon Islands and Papua New Guinea (where vectors differ in terms of habitat requirements), populations at such high risk are expected to be confined to relatively isolated, forested areas with poor health-service access.

Snow *et al.* conclude that the burden of malaria outside Africa is far higher than previous estimates, a prediction that should be broadly verifiable. It would be helpful to have verification of predictions if they are inconsistent with recorded data, such as the 60-times-higher incidence in the WPR, the 0.55 million cases per year among the 6.5 million people in Tajikistan, where transmission is seasonal and *P. vivax* predominates, or the extent to which official figures in Pakistan underestimate incidence by 1,000 times¹. We agree that health information systems miss many malaria cases, particularly among remote populations and

where cases are managed by the private health sector. However, the predictions of Snow *et al.* are undermined by the magnitude of their discrepancy with recent country data based on passive and active case detection provided to the WHO by countries in the WPR.

If the epidemiological characteristics of the WPR are considered, and if similar adjustments are required in other non-African regions, then the global estimates of incidence by Snow *et al.* would be close to current WHO estimates¹⁰. Rather than relying on historical estimates, improved national surveillance is necessary to provide contemporary data on which to base more accurate and verifiable estimates.

David R. Bell, Pernille Jorgensen, Eva Maria Christophel, Kevin L. Palmer
Malaria, Other Vector-borne and Parasitic Diseases Unit, World Health Organization Regional Office for the Western Pacific, PO Box 2932, Manila, Philippines
e-mail: belld@wpro.who.int

1. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. &

Hay, S. I. *Nature* **434**, 214–217 (2005).

2. World Health Organization *Malaria, other vectorborne and parasitic diseases: Regional trend in cases and deaths*, 2003. <http://www.wpro.who.int/sites/mvp/data/> (2005).
3. Lysenko, A. J. & Semashko, I. N. in *Medicinskaja Geografija: Itogi Nauki* 1966 (ed. Lebedew, A. W.) 25–145 (VINITI, Moscow, 1968).
4. Gillies, M. T. in *Malaria: Principles and Practice of Malariology* (eds Wernsdorfer, W. H. & McGregor, I.) 453–485 (Churchill Livingstone, Edinburgh, 1988).
5. United Nations Environment Programme Asia-Pacific Environment Outlook 2: *Regional Resource Centre for Asia and the Pacific* (Pathumthani, Thailand, 2001). www.rrcap.unep.org/reports/apo2.cfm
6. Carter, R., Mendis, K. N. & Roberts, D. *Bull. World Health Organ.* **78**, 1401–1411 (2000).
7. Torres, E. P., Salazar, N. P., Belizario, V. Y. & Saul, A. *Acta Trop.* **63**, 209–220 (1997).
8. Anthony, T. G. *et al. J. Infect. Dis.* **191**, 1558–1564 (2005).
9. Socheat, D. *et al. Southeast Asian J. Trop. Med. Public Health* **34** suppl 4, 1–102 (2003).
10. WHO Expert Committee on Malaria: 20th Report (World Health Organization, Geneva, 2000). <http://www.rbm.who.int/docs/ecr20.pdf>

doi:10.1038/nature04179

Reply: Snow *et al.* reply to this communication (doi:10.1038/nature04180).

MALARIA RISK

Snow *et al.* reply

Reply to: B. L. Nahlen *et al.* (doi:10.1038/nature04178) and D. R. Bell *et al.* (doi:10.1038/nature04179)

We have provided an evidence-based framework for estimating the public-health burden of *P. falciparum* malaria¹ by using a uniform series of structured, transparent and reproducible data and methods to define the cartography of this parasite's clinical impact at a global scale. This contribution to global disease modelling is questioned by Nahlen *et al.*² for United Nations agencies responsible for measuring development targets (see also ref. 3), and our estimation of the malaria burden outside Africa is challenged by Bell *et al.*⁴.

Nahlen *et al.*² point out that the World Health Organization (WHO) estimates the global malaria morbidity burden as being between 300 million and 500 million clinical malaria cases a year⁵. In the single sentence on this subject⁵, we find this estimate represents the combined *Plasmodium falciparum* and *P. vivax* clinical burden and is not sourced. In the following sentence concerning the proportion of global mortality in Africa, a 1993 WHO publication is cited, which does not provide any further citation for how these figures were derived⁶. As with so many quotes of global disease burdens, the empirical basis for this estimate is unclear. The Global Burden of Diseases (GBD) programme of the WHO during the late 1990s sought to replace such guesswork with evidence so that international public-health policy could be better informed⁷. We have therefore chosen to use these 1999 estimates⁸ as the WHO's only global bench-

mark of clinical incidence of *P. falciparum* for which the method was transparent.

In 2004 we provided scientific advice to the WHO's Roll Back Malaria (WHO/RBM) department so that they might begin to improve the estimation of malaria risk⁹. The WHO/RBM chose to use malaria transmission classifications in their original 1968 form¹⁰ and continued to use our modelled, actively detected estimates of malaria morbidity for Africa¹¹; however, they adopted a different approach for outside Africa by triangulating disease reporting and intervention coverage data provided by national governments⁴. By their calculation, there were 311 million clinical cases of *P. falciparum* malaria worldwide in 2002, only about 12% higher than the 273 million estimated by the GBD program⁸, but some 40% lower than our estimate of 515 million cases, which was derived using a more conservative approach to infection risks and active case-detection data that were independent of government statistics¹¹. We therefore cannot agree with Nahlen and colleagues that these two estimates are "consistent"².

The WHO/RBM approach to down-regulating clinical risks according to assumed national reported intervention coverage figures will overcorrect the attributable disease burden when the input incidence estimates are measured under prevailing national-prevention coverage and where we

are uncertain of the subnational estimates of intervention use. Furthermore, the very nature of malaria illness dictates that methodologies institutionally wedded to routine national reporting systems will consistently underestimate malaria burden.

Bell and colleagues⁴ highlight the importance of mapping the extent and intensity of *P. falciparum* transmission at national scales. We agree. Higher-resolution information on forest, land use, human settlement and intervention, and how these relate to dominant vector-species parasite transmission, will provide a more robust framework — with the proviso that this was developed using scientifically defensible models and data. This forms the basis of ongoing work by our group as part of a global Malaria Atlas Project. Unlike the WHO/RBM⁹, we were careful to present only regional comparisons — given the uncertainties in the resolution of the only currently available, global malaria distribution map. For the benefit of the WHO Regional Office for the Western Pacific, the Philippines and Malaysia contributed less than 10% of the clinical burden within the region. As the authors also indicate⁴, Papua New Guinea, the Solomon Islands and Vanuatu have malaria transmission characteristics similar to those in large parts of Africa^{12–14}. As such, we maintain that our delineation of risk in the WPR (western Pacific region) has been conservative.

An epidemiological framework is required to define the spatial extent of malaria risk, how these risks relate to disease burdens, and how these might change by 2010 and 2015 when the WHO/RBM and the United Nations propose to have halved malaria mortality or halted the increasing incidence of malaria, respectively. Defining these targets will demand a credible, science-driven approach to risk mapping. The

use of presumption, best guesses or imperfect data is no longer acceptable.

Robert W. Snow*†, **Carlos A. Guerra**‡, **Abdisalan M. Noor**†, **Hly Y. Myint**§, **Simon I. Hay***‡

*Malaria Public Health & Epidemiology Group, KEMRI/Wellcome Trust Research Laboratories, PO Box 43640, 00100 Nairobi, Kenya
e-mail: rsnow@wtnairobi.mimcom.net

†Centre for Tropical Medicine, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK

‡TALA Research Group, Tinbergen Building, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

§Faculty of Tropical Medicine, Mahidol University, Bangkok 10400, Thailand

1. Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. & Hay, S. I. *Nature* **434**, 214–217 (2005).
2. Nahlen, B. L., Korenromp, E. L., Miller, J. & Shibuya, K. doi:10.1038/nature04178 *Nature* (2005).
3. Attaran, A. *PLoS Med.* (in the press).
4. Bell, D. R., Jorgensen, P., Christophel, E. M. & Palmer, K. L. doi:10.1038/nature04179 *Nature* (2005).
5. WHO Expert Committee on Malaria: 20th Report (World Health Organization, Geneva, 2000). <http://www.rbm.who.int/docs/ecr20.pdf>
6. World Health Organization *A Global Strategy for Malaria Control* (WHO, Geneva, 1993).
7. Murray, C. J. L. & Lopez, A. D. *Lancet* **349**, 1269–1276 (1997).
8. World Health Organization *The World Health Report 1999: Making a Difference — Rolling Back Malaria* (WHO, Geneva, 1999).
9. Hay, S. I., Guerra, C. A. & Snow, R. W. *Determination of Populations at Malaria Risk (MSO/370/19) for Monitoring and Evaluation Group (MERG), Roll Back Malaria (RBM), World Health Organization (WHO)* (WHO, Geneva, 2004).
10. Korenromp, E. L. for the RBM MERG Task Force on Malaria Morbidity: *Malaria Incidence Estimates at Country Level for the Year 2004 — Proposed Estimates and Draft Report* (World Health Organization, Roll Back Malaria, Geneva, 2005). http://mosquito.who.int/docs/incidence_estimations2.pdf.
11. Snow, R. W., Craig, M. H., Newton, C. R. J. C. & Steketee, R. W. *The Public Health Burden of Plasmodium falciparum Malaria in Africa: Deriving the Numbers* (The Disease Control Priorities Project (DCPP) Working Paper Number 11, Washington DC, 2003). www.fic.nigh.gov/dcpp
12. Muller, I., Bockarie, M., Alpers, M. & Smith, T. *Trends Parasitol.* **19**, 253–259 (2003).
13. Over, M., Bakote'e, B., Velayudhan, R., Wilikai, P. & Graves, P. M. *Am. J. Trop. Med. Hyg.* **71**, 214–223 (2004).
14. Maitland, K. et al. *Trans. R. Soc. Trop. Med. Hyg.* **90**, 614–620 (1996).

doi:10.1038/nature04180

Strategies for containing an emerging influenza pandemic in Southeast Asia

Neil M. Ferguson^{1,2}, Derek A.T. Cummings³, Simon Cauchemez⁴, Christophe Fraser¹, Steven Riley⁵, Aronrag Meeyai¹, Sapon Iamsirithaworn⁶ & Donald S. Burke³

Highly pathogenic H5N1 influenza A viruses are now endemic in avian populations in Southeast Asia, and human cases continue to accumulate. Although currently incapable of sustained human-to-human transmission, H5N1 represents a serious pandemic threat owing to the risk of a mutation or reassortment generating a virus with increased transmissibility. Identifying public health interventions that might be able to halt a pandemic in its earliest stages is therefore a priority. Here we use a simulation model of influenza transmission in Southeast Asia to evaluate the potential effectiveness of targeted mass prophylactic use of antiviral drugs as a containment strategy. Other interventions aimed at reducing population contact rates are also examined as reinforcements to an antiviral-based containment policy. We show that elimination of a nascent pandemic may be feasible using a combination of geographically targeted prophylaxis and social distancing measures, if the basic reproduction number of the new virus is below 1.8. We predict that a stockpile of 3 million courses of antiviral drugs should be sufficient for elimination. Policy effectiveness depends critically on how quickly clinical cases are diagnosed and the speed with which antiviral drugs can be distributed.

The continuing spread of H5N1 highly pathogenic avian influenza in wild and domestic poultry in Southeast Asia represents the most serious human pandemic influenza risk for decades^{1,2}. Great potential benefits would be gained from any intervention able to contain the spread of a pandemic strain and eliminate it from the human population. However, the rapid rate of spread of influenza—as witnessed both in annual epidemics and past pandemics^{3–5}—poses a significant challenge to the design of a realistic control strategy.

The basic reproduction number⁶, R_0 , quantifies the transmissibility of any pathogen, which is defined as the average number of secondary cases generated by a typical primary case in an entirely susceptible population. A disease can spread if $R_0 > 1$, but if $R_0 < 1$, chains of transmission will inevitably die out. Hence, the goal of control policies is to reduce R_0 to below 1 by eliminating a proportion $1 - 1/R_0$ of transmission. This can be achieved in three ways: (1) by reducing contact rates in the population (through ‘social distance measures’), (2) by reducing the infectiousness of infected individuals (through treatment or isolation), or (3) by reducing the susceptibility of uninfected individuals (by vaccination or antiviral prophylaxis).

Vaccination and antiviral drugs offer protection against infection and clinical disease. However, although effective vaccines exist for inter-pandemic flu, candidate H5N1 vaccines have unproven effectiveness⁷, and production delays would in any case limit availability in the first months of a pandemic. Antiviral agents—particularly the neuraminidase inhibitors, which show experimental effectiveness against all influenza A subtypes^{8,9}—are therefore a key aspect of recently revised pandemic preparedness plans in several countries¹⁰.

For antivirals to significantly reduce transmission, prophylactic use is necessary. Large-scale prophylaxis has the potential to limit spread substantially in a developed country¹¹, but the very large stocks of drug necessary make this policy impractical if the pandemic is already global. However, might such a policy nonetheless be a feasible strategy

if applied at the source of a new pandemic, when repeated human-to-human transmission is first observed? Here we address this question, and focus on identifying the threshold level of transmissibility below which containment of any new pandemic strain might be feasible.

Modelling pandemic spread

We modelled pandemic spread in Southeast Asia, as this region remains the focus of the ongoing avian H5N1 epidemic and is where most human cases have occurred. Data availability led us to model Thailand rather than any perceived greater risk of emergence compared to other countries in the region; however, we believe our conclusions are also valid for other parts of Southeast Asia.

We constructed a spatially explicit simulation of the 85 million people residing in Thailand and in a 100-km wide zone of contiguous neighbouring countries. The model explicitly incorporates households, schools and workplaces, as these are known to be the primary contexts of influenza transmission^{12–14} (see Fig. 1 and Methods) and because control measures can readily target these locations. Random contacts in the community associated with day-to-day movement and travel were also modelled.

Natural history and transmission parameters

Fundamental to the feasibility of any containment strategy is being able to quantify the transmissibility of the emergent virus, R_0 . Reliable past estimates of transmissibility are rare, perhaps owing to the antigenic diversity of influenza and the consequent complex effect of population immunity on transmission.

We re-analysed incubation period and household transmission data for human influenza (see Methods) and derived new natural history parameters, which predict a profile of infectiousness over time that is remarkably consistent with viral shedding data from experimental infection studies (see Fig. 1g and ref. 15). This profile

¹Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London W2 1PG, UK. ²The Institute of Mathematical Sciences, Imperial College London, 53 Princes Gate, London, SW7 2PG, UK. ³Department of International Health, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, Maryland 21205, USA. ⁴INSERM U707, 27 rue Chaligny, Paris 75571, cedex 12, France. ⁵Department of Community Medicine, 5/F William M. W. Mong Block, Faculty of Medicine Building, The University of Hong Kong, 21 Sassoon Road, Hong Kong. ⁶Bureau of Epidemiology, Department of Diseases Control, Ministry of Public Health, Tivanonda Road, Nonthaburi 11000, Thailand.

gives an estimate of the serial interval or generation time, T_g (the average interval from infection of one individual to when their contacts are infected), of 2.6 days, compared with the value of ~ 4 days assumed by most previous modelling studies¹⁶. Re-analysis of both US and UK 1918 pandemic mortality data using this new value of T_g revises pandemic influenza R_0 estimates⁵ downwards to approximately 1.8 (Fig. 1f and Supplementary Information). This yields a predicted infection attack rate of 50–60% during a pandemic, consistent with what was seen in the first and second waves of past pandemics (see Supplementary Information). An R_0 value of 1.8 is also consistent with annual interpandemic attack rates seen in households where all members were highly susceptible to the prevalent strain¹⁷ (see Supplementary Information). We also assumed that 50% of infections result in clinically recognizable symptoms, with the other 50% being too mild to be diagnosed clinically¹⁸.

We cannot be certain that these parameter estimates would be applicable to any new pandemic strain. It is possible that the mutations or reassortment events that give rise to the new viral strain might initially increase its transmissibility only slightly over the $R_0 = 1$ threshold for self-sustaining transmission. In that case, additional mutations would have to accumulate for viral fitness to increase to its maximum. Given the extended viral shedding (and symptomatic disease) seen in severe human cases of avian H5N1 infection, this might also mean that the T_g of the initial pandemic strain could be considerably greater than for currently circulating human influenza viruses. We therefore examine the ability of control measures to contain pandemic spread not just at a single value of R_0 , but for different values in the range $1 < R_0 < 2$, and analyse model sensitivity to the assumed value of T_g .

Baseline epidemic dynamics

We consider the scenario that a new transmissible ($R_0 > 1$) pandemic strain arises as a result of mutations or a reassortment event in a single individual infected with an avian virus. We seed simulations with a single infection in the most rural third of the population (that is, with the lowest population density), assuming that rural populations are most likely to be exposed to the avian virus. Figure 2 shows the typical pattern of spread for an emergent pandemic initiated by such a seeding event assuming $R_0 = 1.5$, but note that for low R_0 , most epidemics seeded by a single individual go extinct by chance before becoming established in the population.

The pattern of spatial spread (Fig. 2a and Supplementary Video 1) is of interest: for the first 30 days, cases tend to be limited to the region around the seeding location, with few 'sparks' outside that area. However, as case numbers increase exponentially, so does the frequency with which infection events span large distances, and the epidemic rapidly transforms from being predominantly local to country-wide between days 60 and 90 (Fig. 2a–c). Any containment policy needs to be effective before this transition, in part because logistical constraints are likely to preclude containment of a widely disseminated epidemic, but also because the probability of international export of infection becomes high once case numbers reach the thousands (Hollingsworth, D., N. M. F. & Anderson, R. M., unpublished observations).

For $R_0 = 1.5$, the epidemic in the modelled population of 85 million peaks around day 150 and is largely over by day 200 (Fig. 2b), at which point 33% of the population has been infected (Fig. 2d). At $R_0 = 1.8$, the epidemic peaks around day 100 and infects about 50% of the population.

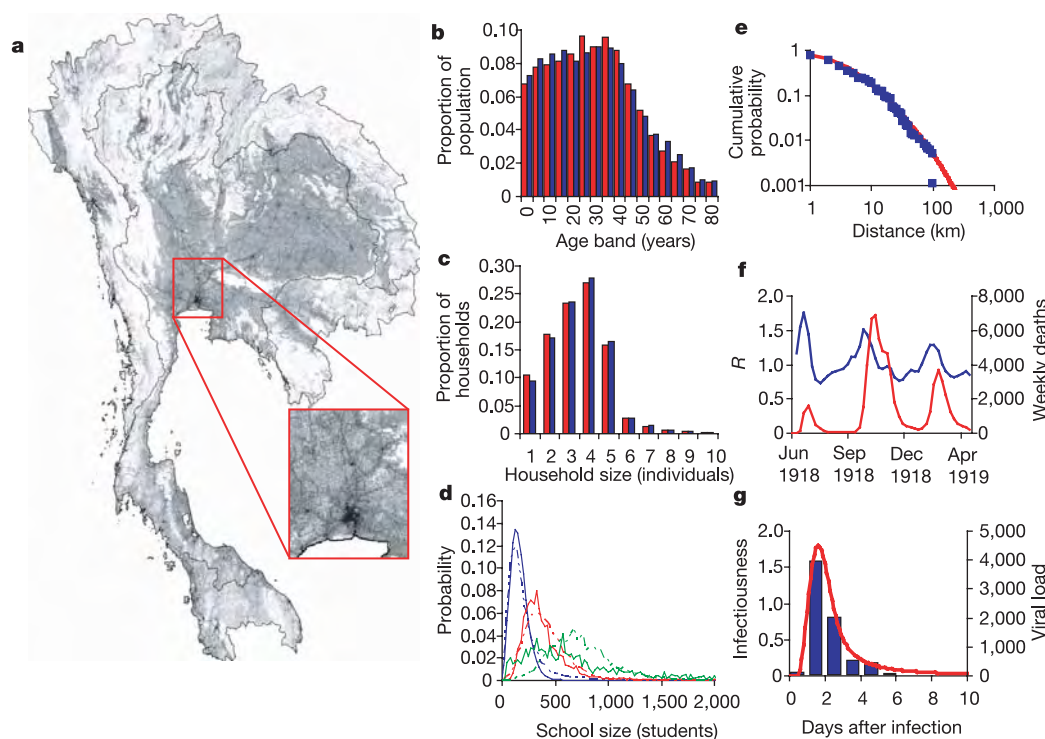


Figure 1 | Data. **a**, Modelled population density of Thailand and 100-km contiguous zone of neighbouring countries, based on Landsat²⁴ data and plotted on a logarithmic scale (light for low density, dark for high density). Inset shows Bangkok in more detail. **b**, Age distribution of Thai population in 2003 in 5-yr bands (blue), and the corresponding age distribution of the simulated population (red). **c**, As **b** but showing distribution of household sizes. **d**, Observed (solid lines) and modelled (dashed lines) distributions of school sizes (blue, elementary; green, secondary; red, mixed). **e**, Probability of travelling over a certain distance to work, estimated from data (blue) and

from the simulated population (red). **f**, Weekly excess influenza-related mortality in 1918–1919 in Great Britain (red), and corresponding estimates of the reproduction number R (blue), calculated assuming $T_g = 2.6$. **g**, Viral shedding data for experimental influenza infection¹⁵ (expressed in tissue culture infective doses (TCID₅₀) per ml of nasal lavage fluid) compared with the modelled profile of infectiousness over time. Note that the infectiousness profile was not fitted to shedding data. See Methods and Supplementary Information for more details.

Effect of antiviral prophylaxis

In evaluating containment strategies, we focus on two principal outcome measures: (1) the probability of preventing a large outbreak (which would eventually lead to a global pandemic), and (2) the number of courses of drug (assumed here to be oseltamivir) required to achieve containment.

Blanket prophylaxis of an entire country or region should be able to eliminate a pandemic virus with an R_0 of 3.6 or greater (see Methods). However, such a policy would require enough drug to prophylax everyone for up to three weeks (that is, at least two courses per person) and is hence unfeasible. Targeted strategies are therefore needed to minimize drug usage while maximizing effect.

Social targeting is the most straightforward approach. This involves prophylaxing individuals in the same household, school or workplace as a newly diagnosed symptomatic case. Unfortunately, if such a policy is only initiated after 20 or more cases, purely social targeting only has a $\geq 90\%$ probability of eliminating the pandemic strain if $R_0 \leq 1.25$ (lowest curve of Fig. 3a; see also Supplementary Information). In reality, at least ten cases might have to be detected to be sure that viral transmissibility had significantly increased¹⁹, and detection and decision-making delays could easily mean 20–30 cases had arisen before policy initiation. A containment policy will therefore probably have to go beyond social targeting in order to succeed. As most community contacts are local, geographic targeting—namely, prophylaxing the whole population in the neighbourhood of the household in which a case is detected—is an obvious policy extension, but one that will no doubt greatly increase the logistical challenges to delivery. In the absence of detailed administrative boundary data, we simulated geographic targeting as the prophylaxis of the population within a ring of a certain radius centred around each detected case, but in practice targeting administrative areas is likely to be more practical. For social or geographic prophylaxis, we assume that individuals are given a single course of ten days of drug, after which time they come off the drug unless more cases have arisen in their vicinity, in which case a second round of prophylaxis is delivered. The policies therefore cease automatically within ten days of the last case being reported.

Our analysis indicates that the additional effort required to deliver a geographic policy pays substantial dividends in terms of policy effectiveness. With a two-day delay from case onset to prophylaxis, a 5-km ring policy is able to contain pandemics with an R_0 of 1.5 (Fig. 3a) at the cost of an average of 2 million courses (Fig. 3b), but the maximum number of courses needed can increase by an (unfeasible) order of magnitude for scenarios in which cases arise in Bangkok at an early stage of the outbreak. Policy effectiveness increases with the radius of the treatment ring selected (but little benefit is gained from exceeding 10 km), as does the number of courses required (Fig. 3b). Policy outcome is still sensitive to the speed of case detection and drug delivery, but containment is always substantially better than for the purely socially targeted policy (Fig. 3d).

As pure radial prophylaxis is costly in terms of drug, we also examined a policy variant that limits the number of people targeted for prophylaxis per case by only targeting the nearest m people (where $m = 10,000$ –50,000) within 10 km of a newly diagnosed case. In areas of low population density, this drug-sparing policy has the same effect as a pure 10-km ring policy, but in high-density areas many fewer courses of drug are used. The improved effectiveness in rural areas outweighs decreased effectiveness in urban areas, resulting in a greater effect than a pure 5-km ring policy and much lower drug use (Fig. 3e, f).

Epidemiologically, elimination occurs either because the treatment strategy reduces R_0 to below 1, or because it reduces R_0 to close to 1 when the epidemic is small, thereby enhancing the probability of random extinction. For scenarios in which the pandemic strain is successfully eliminated, geographic spread is usually limited. For example, the root mean square (r.m.s.) radius of spread is 27 km for $R_0 = 1.5$ using the 5-km radial geographic targeting strategy. When containment is successful, total case numbers are also limited to an average of fewer than 150 cases.

Policies to increase social distance

Measures to increase social distance have been used in past pandemics and remain important options for responding to future pandemics¹. However, predicting the effect of policies such as closing schools and workplaces is difficult, as potentially infectious contacts

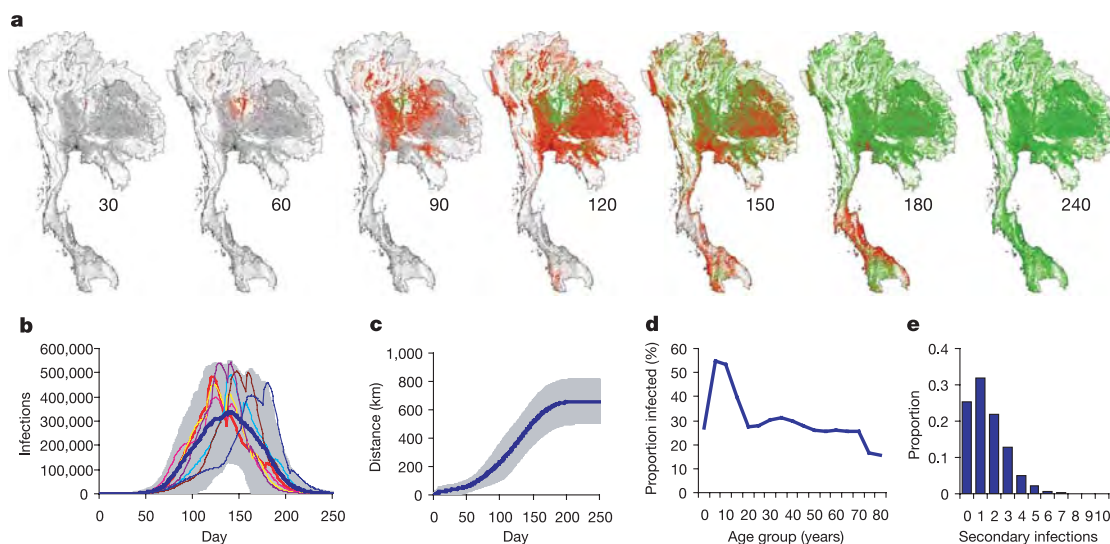


Figure 2 | Expected pattern of spread of an uncontrolled epidemic. **a**, Time sequence (in days) of an epidemic, showing spreading in a single simulation of an epidemic with $R_0 = 1.5$. Red indicates presence of infected individuals, green the density of people who have recovered from infection or died. **b**, Daily incidence of infection over time for $R_0 = 1.5$ in the absence of control measures. Thick blue line represents average for realizations resulting in a large epidemic, grey shading represents 95% confidence limits of the incidence time series. Multiple coloured thin lines show a sample of realizations, illustrating a large degree of stochastic variability. **c**, Root mean square (r.m.s.)

distance from the seed infective for individuals infected since the start of the epidemic as a function of time. Thick blue line represents average distance for realizations resulting in a large epidemic, grey shading represents 95% limits. **d**, Proportion of the population infected by age for $R_0 = 1.5$, averaged across realizations that result in large epidemics. The infection attack rate is 33% for $R_0 = 1.5$ and 50% for $R_0 = 1.8$. **e**, Distribution of the number of secondary cases per primary case during the exponential growth phase of a $R_0 = 1.5$ epidemic. Between 50 and 1,000 realizations were used to calculate all averages (see Supplementary Information).

may be displaced into other settings. Furthermore, it is likely that population contact rates change spontaneously (as well as a result of policy) during severe epidemics (for example, 1918) in response to the perceived risk. Therefore, the estimates of pandemic transmissibility we derive from past pandemics might implicitly incorporate the effects of some degree of social distancing.

We are therefore deliberately conservative in the assumptions made here regarding the effect of school and workplace closure, by assuming that household and random contact rates increase by 100% and 50%, respectively, for individuals no longer able to attend school or work. Figure 4a illustrates how adding area-based school and workplace closure to a drug-sparing prophylaxis policy increases policy effectiveness significantly, with the combined policy having a >90% chance of elimination for $R_0 = 1.7$.

Quarantine zones, in which movements in and out of the affected area are restricted, are another strategy for enhancing containment, and may in any case be thought necessary to prevent population flight from affected areas or deliberate entrance of people into prophylaxis zones to receive drug. Figure 4a (see also Supplementary Video 2) shows that such an area quarantine strategy can greatly increase the effectiveness (to 90% containment at $R_0 = 1.8$) of radial geographic targeted prophylaxis even if only 80% effective at reducing movements. Combining school and workplace closure with area quarantine and prophylaxis further increases policy effectiveness (90% containment at $R_0 = 1.9$), and equally importantly, increases the robustness of the policy to shortcomings in case identification or

treatment rates. For all these policies, containment is typically achieved after fewer than 200 cases have been detected.

Logistical constraints and sensitivity to parameter assumptions

Other constraints may affect the ability of public health authorities to deliver containment policies. Figure 4c shows that the size of an antiviral stockpile can have a substantial effect on policies that use pure radial geographic prophylaxis, as very large numbers of courses are required to prophylax populations around cases arising in large urban areas. However, policies using drug-sparing, geographically targeted prophylaxis (Fig. 4d) retain high effectiveness provided that at least 3 million drug courses are available. For scenarios in which containment fails with a finite stockpile, Fig. 4e shows that even an unsuccessful containment strategy can delay widescale spread by a month or more—a potentially critical window of opportunity for accelerating vaccine production.

Another possible constraint is that capacity to implement these containment policies might not be present in all countries in the region. A policy restricted to one country alone might have a substantially reduced chance of success (Fig. 4f and Supplementary Video 3) should the initial case cluster arise in a border region.

Multiple assumptions are inevitably made when undertaking preparedness modelling for a future emergent infection. Sensitivity analyses are therefore critical for assessing the robustness of policy conclusions. Here, critical assumptions not already discussed include (1) the ratio of within-place to community transmission, (2) the

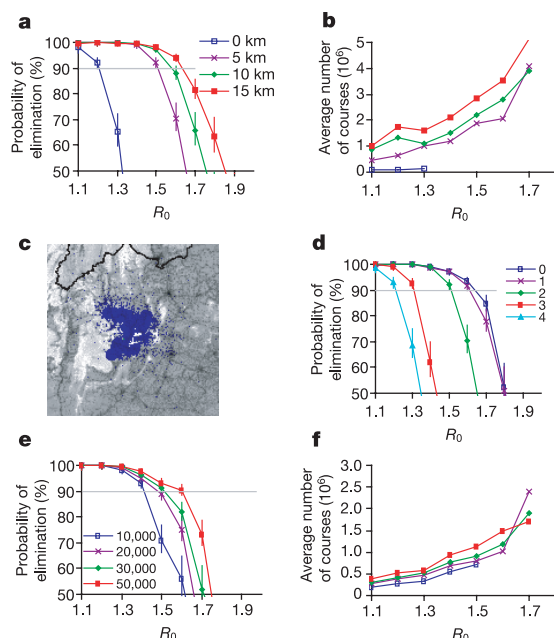


Figure 3 | Prophylaxis strategies. We assume 90% of clinical cases (45% of infections) are detected. Social targeting assumes prophylaxis of 90% of household members and 90% of pupils or colleagues in 90% of the schools or workplaces with detected cases. Geographic targeting assumes 90% of people within 5, 10 or 15 km of a detected case are also prophylaxed. **a**, Probability of eliminating an otherwise large epidemic using social and geographic targeting, as a function of R_0 of the new strain and the radius of prophylaxis. Results assume policy initiation after detection of 20 cases and a two-day delay from case detection to prophylaxis. Error bars show exact 95% confidence limits. **b**, Same as **a**, but showing average number of drug courses required for containment of an otherwise large outbreak. **c**, Map of northern Thailand (150 × 150 km square), showing the extent of spread during one contained $R_0 = 1.8$ epidemic assuming 10-km radial prophylaxis and other parameters as in **a**. Treated areas shown in blue. **d**, Same as **a**, but varying the delay (0–4 days) from case detection to prophylaxis for the 5-km radius policy. **e**, **f**, Same as **a** and **b**, but for drug-sparing policies that target only the nearest 10,000–50,000 people within 10 km of a detected case. Error bars show exact 95% confidence limits.

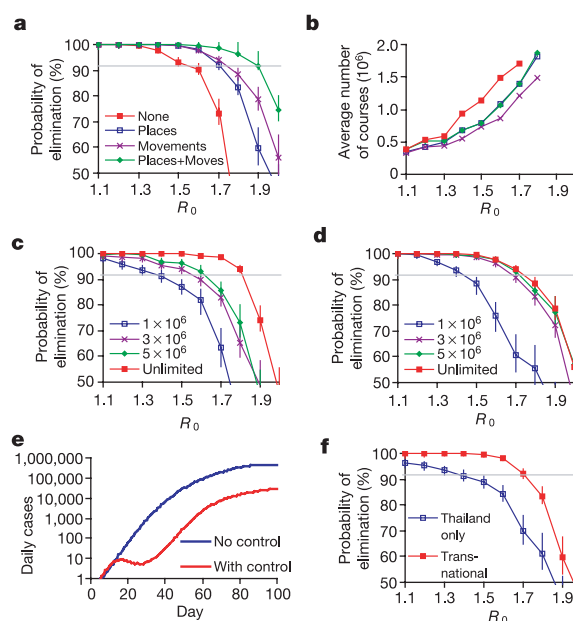


Figure 4 | Social distance measures. **a**, **b**, Same as Fig. 3a, b, but showing the effect of drug-sparing prophylaxis (50,000 courses per case, as Fig. 3e) together with: no social distance measures (red; as Fig. 3); 21-day closure of 90% of schools and 50% of workplaces within 5 km of a detected case (blue); 80% 'area quarantine' (that is, 80% reduction of movement in and out of a zone defined by merging 5-km rings around all detected cases) for 21 days (magenta); or a combination of school/workplace closure and area quarantine (green). **c**, Same as **a** but showing the effect of limiting availability of antiviral drugs to 1, 3 or 5 million courses on the effectiveness of the combined area quarantine and 5-km radial prophylaxis policy. **d**, Same as **c** but for drug-sparing geographic prophylaxis (50,000 courses per case) plus area quarantine. **e**, Case incidence over time without pandemic control measures and with the 3 million course policy of **d**, showing the approximate one-month delay achieved even when containment is unsuccessful ($R_0 = 1.9$). **f**, Same as **a** but showing the reduction in policy effectiveness seen if the combined school/workplace closure and drug-sparing prophylaxis policy is restricted to Thailand alone. Error bars show exact 95% confidence limits.

expected generation time, T_g , of a new pandemic strain (largely determined by the duration of viral shedding and therefore infectiousness), (3) the level of heterogeneity in individual infectiousness (for example, 'superspreaders'²⁰), (4) antiviral efficacy/take-up, and (5) the sensitivity and specificity of case detection during the control programme. The effect of these assumptions on model output is presented in the Supplementary Information. In summary, points (4) and (5) are the most critical, as one might expect. If antiviral coverage or efficacy is considerably less than assumed, then policy effectiveness is substantially reduced. Similarly, if surveillance picks up fewer than 40% of infections (that is, 80% of symptomatic cases), policy effectiveness is again reduced. Poor surveillance specificity (that is, false positives) has an indirect effect on effectiveness as a result of wasted drug and logistical capacity.

Conclusions

We have shown that containment and elimination of an emergent pandemic strain of influenza at the point of origin is feasible using a combination of antiviral prophylaxis and social distance measures. A key conclusion is the need for multiple approaches: simple socially targeted prophylaxis is unlikely to be sufficient if the emergent virus has transmissibility levels near those of previous pandemic viruses. Geographically targeted policies are needed to achieve high levels of containment, with area quarantine being particularly effective at boosting policy effectiveness. The only scenario under which purely socially targeted strategies might be sufficient would be if viral transmissibility evolved incrementally and the emergent virus initially had an R_0 only slightly above 1 (see Supplementary Information); however, R_0 will be probably be uncertain at the time at which containment policies have to be implemented, arguing for precautionary policies that assume transmissibility comparable with that of past pandemics.

A number of key criteria must be met for a high probability of success: (1) rapid identification of the original case cluster, (2) rapid, sensitive case detection and delivery of treatment to targeted groups, preferably within 48 h of a case arising, (3) effective delivery of treatment to a high proportion of the targeted population, preferably >90%, (4) sufficient stockpiles of drug, preferably 3 million or more courses of oseltamivir, (5) population cooperation with the containment strategy and, in particular, any social distance measures introduced, (6) international cooperation in policy development, epidemic surveillance and control strategy implementation. Containment is unlikely if R_0 exceeds 1.8 for the new pandemic strain. Although our analysis of past pandemics suggests that transmissibility will fall below this threshold, it is unlikely that sufficient data will exist to verify this before a containment policy has to be introduced.

The mathematical model we have used to examine the feasibility of pandemic containment is perhaps the largest-scale detailed epidemic micro-simulation yet developed. A key goal of the modelling was parsimony. Although the representation of the population is detailed, this detail is underpinned by available demographic data. The natural history parameters used here have been estimated from primary data on existing influenza strains. The model has five key transmission parameters, of which two were estimated from household data and the remaining three were qualitatively calibrated to historical age-dependent attack rates. We believe that this type of simulation will increasingly become a standard tool for preparedness planning and modelling of new disease outbreaks.

Given the set of criteria listed above for successful containment, the obstacles to practical implementation of such a strategy are undoubtedly formidable. Surveillance is perhaps the single greatest challenge. Success depends on early identification of the first cluster of cases caused by the pandemic strain¹⁹, and on detection of a high proportion of ongoing cases. Some level of mildly symptomatic infection is to be expected (and has been observed for human H5N1 infections²¹), but key to successful containment is the proportion of such cases and their infectiousness. Should the high pathogenicity of

recently reported human infections with the H5N1 virus be even partly maintained, then containment might paradoxically be more likely, as case-ascertainment levels would be higher.

Achieving the rapid delivery of antiviral drugs to a large proportion of the population raises many challenges. Thailand, the country modelled here, is one of the best-prepared and equipped countries in the region in terms of being able to implement a large-scale and very rapid public health intervention. Other countries need considerable development in basic healthcare and disease surveillance infrastructure in order to meet the needs of containment.

Antiviral resistance represents a currently unquantifiable challenge to a prophylaxis-based containment strategy. The key will not be whether genotypic or clinical resistance is seen in a percentage of individuals, but whether resistant viruses are capable of self-sustaining transmission (that is, have $R_0 > 1$). Current evidence indicates that fitness deficits in oseltamivir-resistant strains mean that their transmissibility is limited^{22,23}, but we cannot rule out the possibility that compensatory mutations that increase transmissibility might be selected. If a transmissible resistant strain did emerge during implementation of a containment policy, it would be essential for prophylaxis to cease, lest the wild-type virus be eliminated and the world be left with a pandemic of resistant virus. If prophylaxis were abandoned, the likely higher fitness of the wild-type virus would give every chance for the resistant strain to become excluded from the population.

A feasible strategy for containment of the next influenza pandemic offers the potential to prevent millions of deaths. It is therefore in the interest of all countries to contribute to ensuring that resources, infrastructure and collaborative relationships are in place within the region most likely to be the source of a new pandemic. The challenges are great, but the costs of failure are potentially so catastrophic that it is imperative for the international community to prepare now, to ensure that containment is given the best possible chance of success.

METHODS

Demographic data. The model used Landsat data²⁴ to generate a simulated population realistically distributed across geographic space (Fig. 1a). Thai census data^{25,26} on household size and age distributions were used for demographic parameterization (Fig. 1b, c). Data from the Thai National Statistical Office²⁶ were used to determine the number and proportions of children in school as a function of age, and data from the Thai Department of Education on 24,000 schools (available from the authors upon request) were used to determine the distribution of school sizes (Fig. 1d). Data on travel distances within Thailand were limited; here we used data collected in the 1994 National Migration Survey^{27,28} on distances travelled to work (Fig. 1e and Supplementary Information) to estimate movement kernel parameters. The best-fit kernel had asymptotic power-law form as a function of distance d given by $f(d) \sim 1/[1 + (d/a)^b]$, where $a = 4$ km and $b = 3.8$. Thai workplace sizes²⁹ also follow a power-law distribution³⁰, with an estimated maximum single workplace size of approximately 2,300 and a mean of 21 individuals.

Disease data. The natural history of any H5-based pandemic strain will not be known until it emerges, so we used parameter estimates for current human influenza subtypes, and used sensitivity analyses to investigate what effect deviation from these estimates would have on policy effectiveness (see Supplementary Information). The mean $\pm$ s.d. of the incubation period distribution was estimated as 1.48 ± 0.47 days, on the basis of data from a multiple-exposure event occurring on an aeroplane³¹.

We adopt a more biologically realistic approach than most previous modelling studies (but see ref. 32), and rather than assuming that infectiousness is constant from the end of the latent period until recovery, we model it as a function, $\kappa(T)$ (assumed normalized), depending on the time elapsed from the end of the latent period. The generation time, T_g , is given by the mean latent period plus $\int_0^\infty T\kappa(T)dT$. Experimental infection data³³ indicate the start of symptoms to be coincident with a sharp increase in viral shedding, so we assume that infectiousness starts at the end of the incubation period. We further assume a 0.25-day delay from when symptoms start to when diagnosis or healthcare-seeking behaviour is likely. We used Bayesian methods (see Supplementary Information) to estimate $\kappa(T)$ from data collected in a recent household study of respiratory disease incidence^{34,35}. Combined with the estimated incubation period distribution, this gives the profile of infectiousness shown in Fig. 1g.

T_g is estimated as 2.6 days (95% credible interval: 2.1–3.0), which is shorter than previously assumed (but see ref. 36).

Transmission model. The model is a stochastic, spatially structured, individual-based discrete time simulation. Individuals are co-located in households, with households being constructed to reflect typical generational structure while matching empirical distributions of age structure and household size for Thailand (Fig. 1b, c). Households are randomly distributed in the modelled geographic region, with a local density determined by the Landsat data²⁴. In any time-step of $\Delta T = 0.25$ days, a susceptible individual i has probability $1 - \exp(-\lambda_i \Delta T)$ of being infected, where λ_i is the instantaneous infection risk for individual i . Infection risk comes from 3 sources: (1) household, (2) place, and (3) random contacts in the community. The last of these depends on distance, representing random contacts associated with movements and travel, and is the only means by which infection can cross national borders. Analysis of household infection data (see Supplementary Information), gave a within-household R_0 of 0.6 and an overall R_0 of 1.8. We partition non-household transmission to give levels of within-place transmission comparable with household transmission (that is, $R_0 \approx 0.6$) and to qualitatively match 1957 influenza pandemic age-specific attack rates. When varying R_0 , the relative proportions of household, place and community transmission were kept fixed. Full model details are given in the Supplementary Information.

Antiviral drug action. We use recent statistically rigorous estimates of antiviral efficacy³⁷, but these are broadly consistent with previous estimates²². Prophylaxis of uninfected individuals is assumed to reduce susceptibility to infection by 30%, reduce infectiousness if infection occurs by 60%, and reduce the probability of clinically recognizable symptoms by 65% (ref. 37). In theory, blanket prophylaxis of a population should be able to contain a pandemic with an R_0 of $1/[(1 - 0.6)(1 - 0.3)]$, or approximately 3.6. Treatment of a symptomatic case is assumed to reduce infectiousness by 60% from when treatment is initiated. Overall, for the parameter values used here, antiviral treatment of a symptomatic case can reduce total infectiousness throughout the course of infection by a maximum of 28%.

Received 2 June; accepted 14 July 2005.

Published online 3 August 2005.

- World Health Organisation. Avian influenza: assessing the pandemic threat. (http://www.who.int/csr/disease/influenza/WHO_CDS_2005_29/en/index.html) (2005).
- Abbott, A. & Pearson, H. Fear of human pandemic grows as bird flu sweeps through Asia. *Nature* **427**, 472–473 (2004).
- Spicer, C. C. & Lawrence, C. J. Epidemic influenza in Greater London. *J. Hyg. (Lond.)* **93**, 105–112 (1984).
- Fleming, D. M., Zambon, M., Bartelds, A. I. M. & de Jong, J. C. The duration and magnitude of influenza epidemics: A study of surveillance data from sentinel general practices in England, Wales and the Netherlands. *Eur. J. Epidemiol.* **15**, 467–473 (1999).
- Mills, C. E., Robins, J. M. & Lipsitch, M. Transmissibility of 1918 pandemic influenza. *Nature* **432**, 904–906 (2004).
- Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1992).
- Stohr, K. & Esveld, M. Will vaccines be available for the next influenza pandemic? *Science* **306**, 2195–2196 (2004).
- Leneva, I. A., Roberts, N., Govorkova, E. A., Goloubeva, O. G. & Webster, R. G. The neuraminidase inhibitor GS4104 (oseltamivir phosphate) is efficacious against A/Hong Kong/156/97 (H5N1) and A/Hong Kong/1074/99 (H9N2) influenza viruses. *Antiviral Res.* **48**, 101–115 (2000).
- Brooks, M. J., Sasadeusz, J. J. & Tannock, G. A. Antiviral chemotherapeutic agents against respiratory viruses: where are we now and what's in the pipeline? *Curr. Opin. Pulm. Med.* **10**, 197–203 (2004).
- Coombes, R. UK stocks up on antiviral drug to tackle flu outbreak. *Br. Med. J.* **330**, 495 (2005).
- Longini, I. M. Jr, Halloran, M. E., Nizam, A. & Yang, Y. Containing pandemic influenza with antiviral agents. *Am. J. Epidemiol.* **159**, 623–633 (2004).
- Viboud, C. et al. Risk factors of influenza transmission in households. *Br. J. Gen. Pract.* **54**, 684–689 (2004).
- Principi, N., Esposito, S., Gasparini, R., Marchisio, P. & Crovari, P. Burden of influenza in healthy children and their households. *Arch. Dis. Child.* **89**, 1002–1007 (2004).
- Heymann, A., Chodick, G., Reichman, B., Kokia, E. & Laufer, J. Influence of school closure on the incidence of viral respiratory diseases among children and on health care utilization. *Pediatr. Infect. Dis. J.* **23**, 675–677 (2004).
- Hayden, F. G. et al. Local and systemic cytokine responses during experimental human influenza A virus infection. Relation to symptom formation and host defense. *J. Clin. Invest.* **101**, 643–649 (1998).
- Elveback, L. R. et al. An influenza simulation model for immunization studies. *Am. J. Epidemiol.* **103**, 152–165 (1976).
- Longini, I. M. Jr, Koopman, J. S., Haber, M. & Cotsonis, G. A. Statistical inference for infectious diseases. Risk-specific household and community transmission parameters. *Am. J. Epidemiol.* **128**, 845–859 (1988).
- Fox, J. P., Hall, C. E., Cooney, M. K. & Foy, H. M. Influenzavirus infections in Seattle families, 1975–1979. I. Study design, methods and the occurrence of infections by time and age. *Am. J. Epidemiol.* **116**, 212–227 (1982).
- Ferguson, N. M., Fraser, C., Donnelly, C. A., Ghani, A. C. & Anderson, R. M. Public health risk from the avian H5N1 influenza epidemic. *Science* **304**, 968–969 (2004).
- Riley, S. et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science* **300**, 1961–1966 (2003).
- World Health Organisation. WHO inter-country consultation influenza A/H5N1 in humans in Asia. http://www.who.int/csr/disease/avian_influenza/H5N1%20Intercountry%20Assessment%20final.pdf (2005).
- Ferguson, N. M., Mallett, S., Jackson, H., Roberts, N. & Ward, P. A population-dynamic model for evaluating the potential spread of drug-resistant influenza virus infections during community-based use of antivirals. *J. Antimicrob. Chemother.* **51**, 977–990 (2003).
- Herlocher, M. L. et al. Influenza viruses resistant to the antiviral drug oseltamivir: transmission studies in ferrets. *J. Infect. Dis.* **190**, 1627–1630 (2004).
- Oakridge National Laboratory. Landsat global population data. <http://www.ornl.gov/sci/gist/landsat> (2003).
- National Statistical Office Thailand. Population Census 2000. http://web.nso.go.th/pop2000/pop_e2000.htm (2000).
- National Statistical Office Thailand. Population and Labor Statistics. <http://web.nso.go.th/eng/stat/subject/subject.htm#cata1> (1999).
- National Statistical Office Thailand. National Migration Survey. <http://web.nso.go.th/eng/stat/migrant/migrant.htm> and <http://opr.princeton.edu/archive/nmst/> (1994).
- Chamrathirong, A. et al. *National Migration Survey of Thailand* (Institute for Population and Social Research, Mahidol Univ., Bangkok, 1995).
- Ministry of Labour Thailand. Yearbook of Labour Statistics. <http://www1.mol.go.th/download/chap031-2.pdf> (2003).
- Axtell, R. L. Zipf distribution of U.S. firm sizes. *Science* **293**, 1818–1820 (2001).
- Moser, M. R. et al. An outbreak of influenza aboard a commercial airliner. *Am. J. Epidemiol.* **110**, 1–6 (1979).
- Fraser, C., Riley, S., Anderson, R. M. & Ferguson, N. M. Factors that make an infectious disease outbreak controllable. *Proc. Natl Acad. Sci. USA* **101**, 6146–6151 (2004).
- Hayden, F. G. et al. Use of the oral neuraminidase inhibitor Oseltamivir in experimental human influenza. *J. Am. Med. Assoc.* **282**, 1240–1246 (1999).
- Carrat, F. et al. Influenza burden of illness: estimates from a national prospective survey of household contacts in France. *Arch. Intern. Med.* **162**, 1842–1848 (2002).
- Cauchemez, S., Carrat, F., Viboud, C., Valleron, A. J. & Boelle, P. Y. A Bayesian MCMC approach to study transmission of influenza: application to household longitudinal data. *Stat. Med.* **23**, 3469–3487 (2004).
- Flahault, A. et al. Modelling the 1985 influenza epidemic in France. *Stat. Med.* **7**, 1147–1155 (1988).
- Yang, Y., Longini, I. M. & Halloran, M. E. Design and evaluation of prophylactic interventions using infectious disease incidence data from close contact groups. Department of Biostatistics, Emory University, Technical Report 04-09 http://www.sph.emory.edu/bios/tech/Tech_Report_04-09.pdf (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the National Institute of General Medical Sciences MIDAS Program (N.M.F., D.A.T.C. and D.S.B.), the Medical Research Council (N.M.F.), the Royal Society (N.M.F. and C.F.), the Howard Hughes Medical Institute (N.M.F.), the Research Fund for the Control of Infectious Diseases of the Hong Kong SAR government (S.R.) and INSERM (S.C.) for research funding. We thank F. Carrat for providing household data used in this study, and N. Cox, F. Hayden, B. Schwartz, K. Stohr and members of the MIDAS consortium for useful discussions. We thank the MIDAS informatics group for computational resources.

Author Contributions N.M.F. designed, implemented and ran the model, integrated the demographic and disease datasets used, and drafted and revised the text. All other authors edited or commented on the text. D.A.T.C. identified, collated and processed the Thai demographic and travel datasets used and provided input on model assumptions. S.C. analysed the French household dataset used to estimate key epidemiological parameters. C.F. performed analytical modelling of household transmission, which aided the verification of the simulation and gave suggestions on control strategies to be modelled. S.R. contributed to the design of some algorithms within the simulation model. A.M. assisted with collating data on Thai schools and administrative boundaries. S.I. provided feedback on the realism of model assumptions and the likely feasibility of different control options. D.S.B. provided input into model design and assumptions, advised on the presentation of results, assisted with data collection and organized meetings with stakeholders.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to N.M.F. (neil.ferguson@imperial.ac.uk).

Crystal structure of a bacterial homologue of Na^+/Cl^- -dependent neurotransmitter transporters

Atsuko Yamashita¹, Satinder K. Singh¹, Toshimitsu Kawate¹, Yan Jin² & Eric Gouaux^{1,2}

Na^+/Cl^- -dependent transporters terminate synaptic transmission by using electrochemical gradients to drive the uptake of neurotransmitters, including the biogenic amines, from the synapse to the cytoplasm of neurons and glia. These transporters are the targets of therapeutic and illicit compounds, and their dysfunction has been implicated in multiple diseases of the nervous system. Here we present the crystal structure of a bacterial homologue of these transporters from *Aquifex aeolicus*, in complex with its substrate, leucine, and two sodium ions. The protein core consists of the first ten of twelve transmembrane segments, with segments 1–5 related to 6–10 by a pseudo-two-fold axis in the membrane plane. Leucine and the sodium ions are bound within the protein core, halfway across the membrane bilayer, in an occluded site devoid of water. The leucine and ion binding sites are defined by partially unwound transmembrane helices, with main-chain atoms and helix dipoles having key roles in substrate and ion binding. The structure reveals the architecture of this important class of transporter, illuminates the determinants of substrate binding and ion selectivity, and defines the external and internal gates.

Communication between neurons in the brain is critically dependent on the transmission of nerve impulses through chemical synapses, junctions at which electrical signals are relayed from one neuron to another via classical¹ neurotransmitters such as serotonin, norepinephrine, dopamine, glycine and GABA (γ -aminobutyric acid). Arrival of an action potential at the presynaptic terminal triggers release of these small molecules, which can then activate either millisecond-acting, ligand-gated ion channels or more slowly acting G-protein-coupled receptors¹. After release and receptor activation, these neurotransmitters are actively cleared from the synapse primarily by specific, high-affinity integral membrane transporters located on the presynaptic terminal and surrounding glial cells. This energy-driven re-uptake not only returns synaptic neurotransmitter concentrations to basal levels but also allows for the replenishment of neurotransmitter stores in the lumen of presynaptic secretory vesicles by a separate class of vesicular carriers².

Na^+/Cl^- -dependent transporters, also referred to as neurotransmitter sodium symporters (NSS), use sodium and chloride electrochemical gradients to catalyse the thermodynamically uphill movement of a wide array of substrates, including the biogenic amines (serotonin, dopamine, norepinephrine), amino acids (GABA, glycine, proline, taurine) and osmolytes (betaine, creatine)². NSS members are distinct, in terms of amino acid sequence and functional properties, from the high-affinity, sodium-dependent glutamate transporters—the second major family of plasma membrane neurotransmitter transporters². Dysfunction of Na^+/Cl^- -dependent transporters contributes to multiple disorders, including depression³, Parkinson's disease³, orthostatic intolerance³ and epilepsy⁴. Notably, they are the target of addictive substances such as cocaine and amphetamine⁵ as well as therapeutic agents such as anticonvulsants⁶ and the selective serotonergic re-uptake inhibitors⁷.

A breakthrough in the structure/function analysis of Na^+/Cl^- -dependent transporters occurred in the early 1990s, when the gene

encoding the rat GABA transporter type 1 (GAT1)⁸, along with other family members^{9,10}, was cloned. Analysis of the predicted amino acid sequences provided evidence for 12 transmembrane segments (TMs), with the amino and carboxy termini residing in the cytoplasm—predictions that were later confirmed by experiment¹¹. Subsequent studies implicated residues in inhibitor recognition, substrate specificity, gating and ion binding. Crucial residues include a strictly conserved tyrosine in TM3 that is indispensable for substrate binding and transport^{12,13}, a position in TM1 occupied by either an aspartate or glycine, partially responsible for distinguishing between monoamine and amino acid substrates, respectively¹⁰, and a negatively charged residue in extracellular loop 5 (EL5)/TM10 that is postulated to comprise part of an external gate¹⁴ (Fig. 1a). Sodium:chloride:substrate stoichiometries have also been elaborated and vary from 1:1:1, 2:1:1 to 3:1:1 among different transporters^{15–17}. At the present time, however, there is no structural paradigm for the interpretation of this wealth of functional data. Furthermore, there is no information, at the level of atomic detail, on the molecular principles underlying sodium-ion selectivity, sodium ion–substrate coupling and transport mechanism.

To obtain atomic resolution structural data on an NSS family member, we exploited the fact that prokaryotic organisms harbour NSS homologues^{9,18}. Clusters of high sequence conservation, including functionally important residues, are distributed throughout the primary structure, although the overall sequence identity between the eukaryotic and prokaryotic counterparts is only 20–25% (Fig. 1a). Here we present the crystal structure of a bacterial homologue from *Aquifex aeolicus* (LeuT_{AA}) and the molecular insights it unveils about the family of Na^+/Cl^- -dependent neurotransmitter transporters.

Structure determination

LeuT_{AA} yielded crystals readily, and initial phases were determined by multiwavelength anomalous dispersion (MAD)¹⁹ using a single

¹Department of Biochemistry and Molecular Biophysics and ²Howard Hughes Medical Institute, Columbia University, 650 West 168th Street, New York, New York 10032, USA.

crystal grown from selenomethionine-labelled protein and diffraction data measured to Bragg spacings of 1.9 Å (Supplementary Table 1). The resultant electron density map was unambiguous over almost the entire molecule. An initial model derived from automatic tracing was subjected to rounds of manual rebuilding and crystallographic refinement using a native data set to 1.65 Å resolution. The final model contains residues 5–133 and 135–513 of LeuT_{Aa}, leucine, two sodium ions, a chloride ion, five detergent molecules and 210 water molecules (Table 1).

Transporter architecture

The LeuT_{Aa} protomer resembles a shallow 'shot glass', with the opening facing the extracellular space, the base facing the cytoplasm, and the bottom of the 'glass' located ~6 Å into the bilayer-spanning portion of the transporter (Fig. 2a, b). In the current structure there

are no solvent-accessible channels traversing the membrane-spanning portion of the protein. The protomer is ~70 Å tall and ~48 Å in diameter and comprises 12 transmembrane helical regions (TM1 to TM12), together with numerous loops and helices on the intracellular and extracellular surfaces. To the best of our knowledge, the fold of LeuT_{Aa} is not similar to that of any previously reported membrane protein structure.

There is an unanticipated internal structural repeat in the first ten transmembrane helices of the LeuT_{Aa} protomer, relating TM1–TM5 and TM6–TM10 by a pseudo-two-fold axis located in the plane of the membrane (Figs 1b and 2c). Specifically, the α-carbon atoms of the 130 residues comprising helices TM1–TM5 can be superposed onto those from TM6–TM10 by a rotation of 176.5°, yielding a root mean square deviation (r.m.s.d.) of 5.3 Å. This approximate symmetry is not detectable in the amino acid sequence, and it positions the

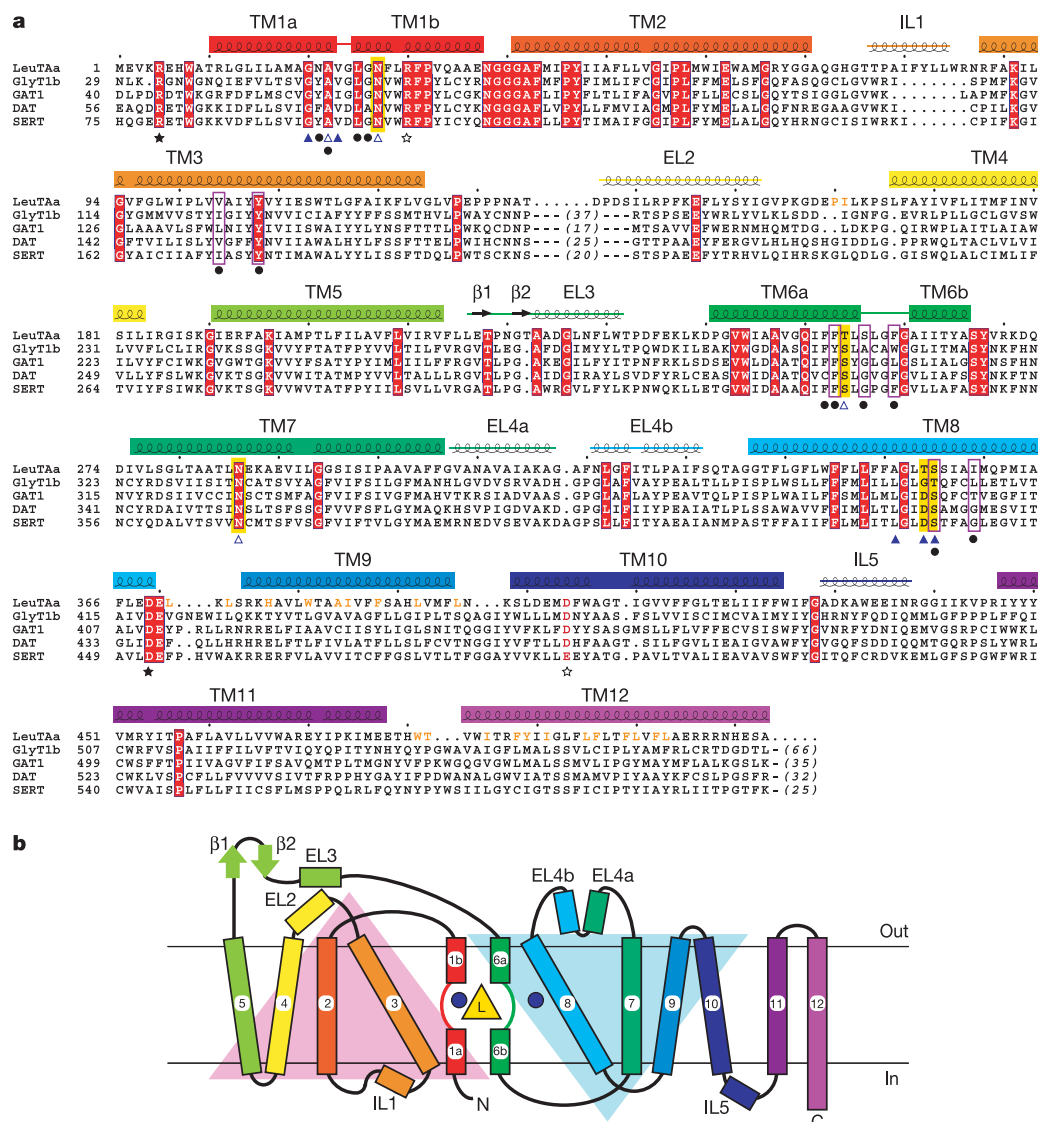


Figure 1 | Amino acid sequence alignment and secondary structure of LeuT_{Aa}. **a**, Amino acid sequence alignment of *A. aeolicus* LeuT_{Aa} (NP_214423) with the human transporter homologues for glycine (GlyT1b; I57956), GABA (GAT1; P30531), dopamine (DAT; Q01959) and serotonin (SERT; P31645) using Psi-BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/>) with manual adjustment. Strictly conserved residues are highlighted in red, and α-helices and β-strands in LeuT_{Aa} are depicted as coils and arrows, respectively. The open and filled blue triangles show residues involved in coordinating sodium ions Na1 and Na2, respectively, and the residues whose side-chain atoms interact with the sodium ions are further highlighted in

yellow. The filled black circles indicate the residues involved in L-leucine binding, and the residues whose side chains interact with the leucine are enclosed by purple boxes. The open and filled stars indicate the charged pairs at the extracellular and cytoplasmic entrances, respectively. The residues in the LeuT_{Aa} dimer interface are shown in orange letters. For the eukaryotic transporters, residues at the N- and C termini and between TM3 and TM4 are truncated in the alignment, and the numbers of truncated residues are shown in parentheses. **b**, The LeuT_{Aa} topology. The positions of leucine and the two sodium ions are depicted as a yellow triangle and blue circles, respectively.

Table 1 | Refinement statistics

Resolution (Å)	50–1.65
$R_{\text{work}}/R_{\text{free}}^*$	0.199/0.217
Number of atoms	
Protein	4,044
Substrate/ion	12
Detergent	100
Water	210
B-factors	
Protein	27.7
Substrate/ion	19.2
Detergent	58.8
Water	44.2
R.m.s.d.	
Bond lengths (Å)	0.005
Bond angles (°)	1.11

* $R_{\text{work}} = \sum ||F_o| - |F_c|| / \sum |F_o|$. R_{free} is the R -value for a subset of 5% of the reflection data, which were not included in the crystallographic refinement.

pseudo repeats with opposite orientations in the membrane bilayer. Not surprisingly, there are bacterial homologues that possess only the first 10–11 transmembrane helices¹⁸, thus reinforcing the concept that the first ten helices form the essential core of Na^+/Cl^- -dependent neurotransmitter transporters. Interestingly, other membrane transport proteins, such as aquaporin²⁰ and the oxalate transporter²¹, possess pseudo-two-fold axes parallel and perpendicular to the membrane.

TM1 and TM6, which harbour the greatest density of conserved residues (Fig. 1a), are oriented antiparallel to one another and, most importantly, are not continuous helices. In fact, TM1 and TM6 have breaks in helical structure at positions approximately halfway across the membrane bilayer. In TM1, residues Val 23 and Gly 24 adopt an extended conformation, linking segments 1a and 1b. The interruption of helical structure is greater in TM6, with residues Ser 256 to Gly 260 adopting extended, non-helical conformations. These two

breaks in helical structure expose main-chain carbonyl oxygen and nitrogen atoms for hydrogen bonding and ion coordination. TM3 and TM8, long helices tilted by $\sim 50^\circ$ from the membrane normal, are also related by the pseudo-two-fold axis and possess key conserved amino acid residues located near the unwound segments of TM1 and TM6. As discussed below, TM3, TM8 and the zone surrounding the TM1 and TM6 unwound regions comprise the substrate and sodium ion binding sites.

The remaining transmembrane segments surround TM1, TM3, TM6 and TM8, support the protein core and comprise most of the lipid-exposed LeuT_{Aa} surface, with TM2 and TM7 buttressing TM6 and TM1, respectively. A prominent V-shaped structure formed by TM4–TM5 and a pseudo-two-fold related inverted V created by TM9–TM10 hold TM3 and TM8 like pincers. TM10 is followed by TM11 and TM12, flanking the outer surfaces of TM9 and TM10. TM12 is a helix to the C terminus of the protein and protrudes the deepest into the cytoplasm (Fig. 2a). Conserved amino acid residues are observed throughout TM2, TM4, TM5, TM7 and TM9–TM12 (Fig. 1a), reinforcing their importance in structure and function.

There are substantial helical and loop segments on the solvent-exposed surfaces of LeuT_{Aa} , some of which are situated in regions that suggest a role in the transport mechanism. The first six residues at the N terminus (Arg 5 to Thr 10) possess an extended structure and are located in the crevices between the ends of TM1a, TM4, TM5, TM6b and TM8 (Supplementary Fig. 1). Specifically, two strictly conserved residues, Arg 5 and Trp 8, make numerous interactions in these crevices. Between TM2 and TM3 are a re-entrant segment of polypeptide and a short helix (IL1) that occlude the intracellular face of LeuT_{Aa} (Supplementary Fig. 1). On the opposite side of the bilayer, between TM3 and TM4, a stretch of 41 residues (EL2) reaches 'up' from TM3 via extended polypeptide and then returns 'down' to TM4 via a helix and a β -turn. In eukaryotic transporters the number of residues between TM3 and TM4 is substantially greater (Fig. 1a), harbouring glycosylation sites⁹ and a probable disulphide bond^{22,23}.

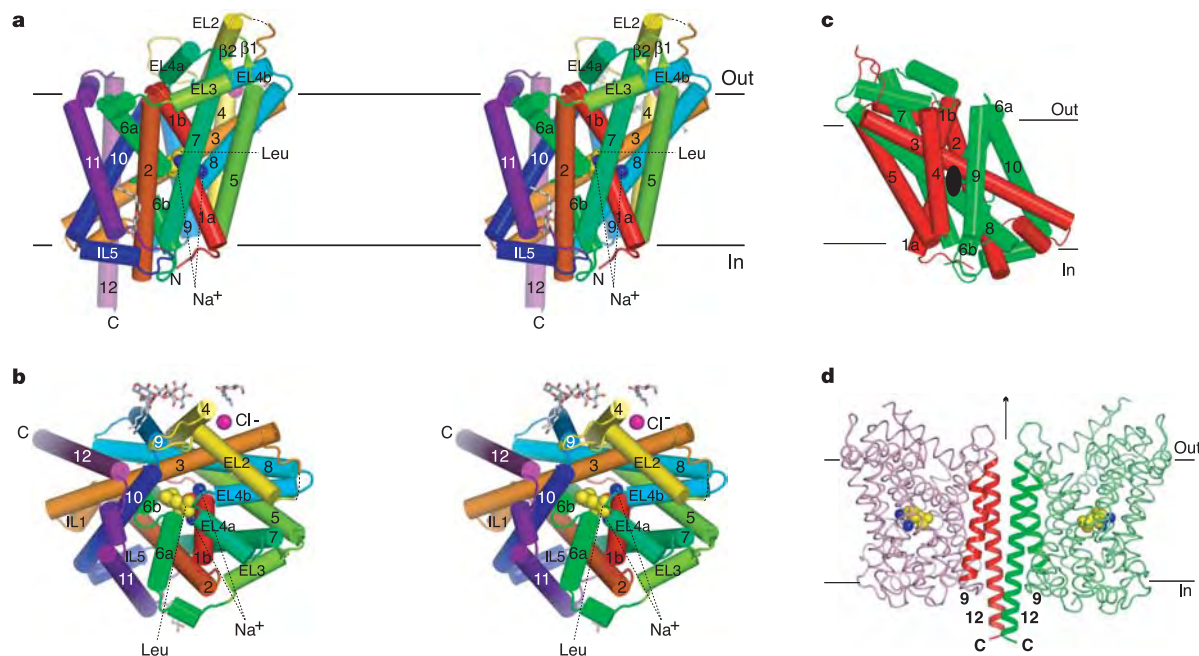


Figure 2 | LeuT_{Aa} structure. **a**, Stereoview in the plane of the membrane. Bound L-leucine, two sodium ions and a chloride ion are shown as CPK models in yellow, blue and magenta, respectively. Detergent molecules (β -OG) are shown as stick models. Colouring and numbering of helices (cylinders) and strands (arrows) are the same as in Fig. 1. **b**, Stereoview from the extracellular side. **c**, Pseudo-symmetry in the LeuT_{Aa} molecule. Shown are transmembrane segments TM1–TM10, viewed approximately parallel to

the pseudo-two-fold axis of symmetry that relates TM1–TM5 (red) and TM6–TM10 (green). The pseudo-two-fold rotation axis is depicted as a black ellipsoid. **d**, The LeuT_{Aa} dimer, with the subunits related by the crystallographic two-fold axis of symmetry (black arrow). One protomer is light green, the other is pink, and TM9 and TM12 are highlighted and numbered. Leucine and the sodium ions are shown in CPK representation in yellow and blue, respectively.

The helical segment of EL2 is juxtaposed across from EL4, a segment of polypeptide between TM7 and TM8 that is composed of two short helices separated by an acute bend. The EL4a helix, in turn, makes numerous contacts with TM1b. Together, EL2 and EL4 define a portion of the shot-glass rim above the membrane bilayer.

LeuT_{Aa} forms a dimer in the crystal, located on the two-fold axis of crystallographic symmetry, and the two protomers adopt a parallel orientation (Fig. 2d). The dimer interface is formed by EL2, TM9 and TM12, and the latter two elements of structure, together with their symmetry partners, form a four-helix bundle (Fig. 1a; see also Supplementary Fig. 2). The relevance of this interface to eukaryotic NSS members, most of which are thought to form oligomers, is supported by experiments on SERT, which indicate that TM12, together with TM11, participate in multimerization²⁴. TM2, TM4 and TM6 have also been implicated in dimerization of some eukaryotic counterparts²⁵, but because TM2 and TM6 are substantially sequestered within the LeuT_{Aa} core, it is unlikely that these transmembrane segments participate in extensive protomer-protomer contacts in eukaryotic transporters. However, TM4, which lies on the surface, could partake in multimerization, although it is not part of the LeuT_{Aa} dimer interface.

Leucine binding site

Prominent non-protein electron density located in the core of the LeuT_{Aa} protein, near the middle regions of TM3 and TM8 and adjacent to the unwound regions of TM1 and TM6 (Fig. 3a),

pin-pointed the location of substrate and sodium ion binding sites. The larger electron density feature indicated that it was L-leucine, whereas the other two were interpreted as sodium ions (see below). Indeed, LeuT_{Aa} reconstituted into liposomes exhibited sodium-dependent uptake of L-leucine (Fig. 3b). Transport was not chloride dependent, consistent with the absence of a chloride ion located near the leucine and sodium ions, although a chloride ion is bound on the outer surface of the protein (Fig. 2a, b; see also Supplementary Fig. 3). Leucine uptake by LeuT_{Aa} proteoliposomes was stimulated ~20–30% by valinomycin, suggesting that under the current assay conditions leucine transport is enhanced by vesicles with an internal negative potential, and therefore transport may be electrogenic (Supplementary Fig. 4a). Further experiments, however, are required to establish the stoichiometries of substrate and ion(s) in the transport cycle of LeuT_{Aa}.

The most striking feature of the leucine binding site is that exposed main chain atoms from the unwound regions of TM1 and TM6 make most of the contacts with the α -amino and α -carboxy groups of the bound leucine (Fig. 3c). Unwinding of TM1 and TM6 not only affords direct hydrogen-bonding partners but also allows the α -amino and α -carboxy groups to bind close to the ends of the helical segments and exploit α -helix dipole moments. Specifically, the negatively charged ends of the TM1a and TM6a helix dipoles are proximal to the amino group of leucine, whereas the positively charged end of the TM1b helix dipole is near the carboxy group. Notably, no residues that bear a formal charge interact with the α

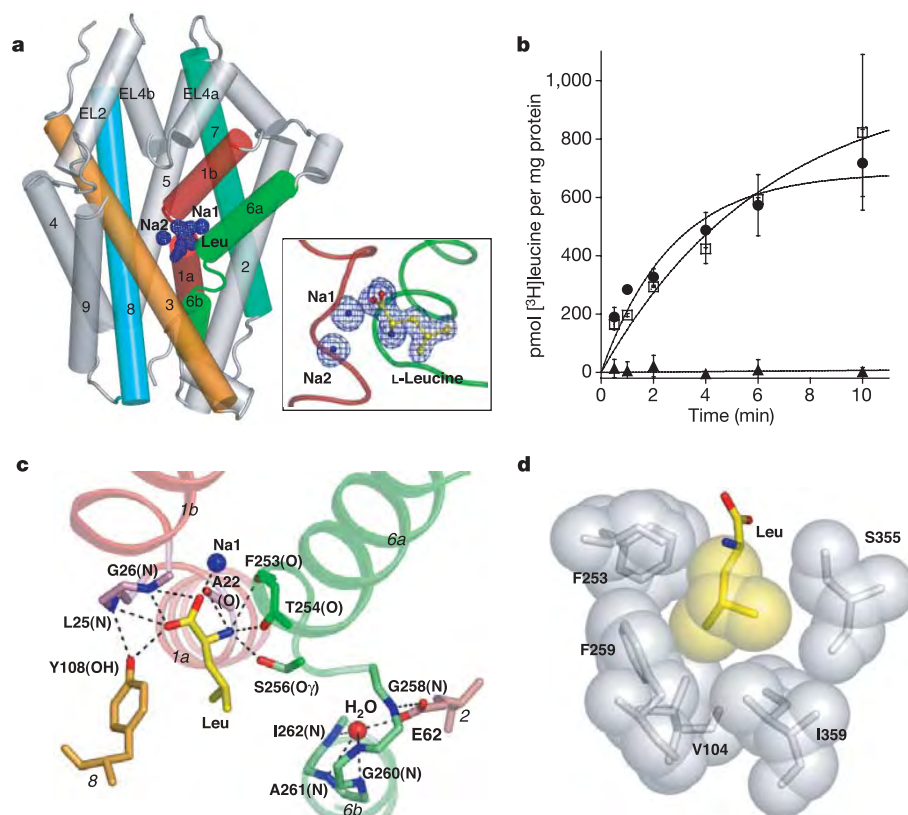


Figure 3 | Leucine binding site. **a**, $F_o - F_c$ simulated-annealing omit map, contoured at 3.8σ , showing density for leucine and sodium ions, where leucine and the two sodium ions were omitted from the simulated annealing run and subsequent phase calculation. View of LeuT_{Aa} rotated 110° anticlockwise in the membrane plane and 25° towards the reader relative to the view in Fig. 2a. TM10–TM12 are omitted from the figure for clarity. A magnified view of the density is shown in the enclosed box. **b**, [³H]Leucine transport by LeuT_{Aa} reconstituted into liposomes. The filled circles, open squares and filled triangles represent uptake from an external solution

containing NaCl (sodium and chloride gradients), sodium gluconate (sodium gradient) and KCl (chloride gradient), respectively. The data points were fit to a single exponential (solid line). Error bars for each data point represent the standard deviation for duplicate measurements.

c, Hydrogen bonds and ionic interactions in the leucine binding pocket are depicted as dashed lines. **d**, Hydrophobic interactions between the leucine and LeuT_{Aa}. Van der Waals surfaces for the leucine side chain and interacting residues are shown as spheres. Tyr 108 and Ser 256 are omitted from the figure for clarity.

substituents of leucine. Furthermore, the leucine binding site is completely dehydrated, and the closest water molecules are $>9\text{ \AA}$ away from the leucine α -carbon atom (Supplementary Fig. 5). Although we cannot directly discern the ionization state of the α substituents from our electron density maps, the pattern of hydrogen bonding interactions suggests that leucine is bound as a zwitterion.

Copious polar contacts, therefore, satisfy the hydrogen bonding requirements of the charged α substituents. The amino group is coordinated by main-chain carbonyl oxygens from Ala 22 in TM1, Phe 253 and Thr 254 in TM6, and by a side-chain hydroxyl from Ser 256 in TM6. The carboxy group interacts with a sodium ion, amide nitrogens from Leu 25 and Gly 26 in TM1, and a hydroxyl from Tyr 108 in TM3 (Fig. 3c). The Tyr 108 hydroxyl also forms a hydrogen bond with the main-chain amide nitrogen of Leu 25, an interaction that may function as a latch to stabilize the irregular structure near the unwound region in TM1 (Fig. 3c). This tyrosine is strictly conserved among all NSS family members (Fig. 1a) and has been implicated in substrate binding and transport in GAT1 (ref. 12), SERT¹³ and GLYT2a²⁶. The structure of the unwound region in TM6 is stabilized by Glu 62, which interacts with the amide nitrogen of Gly 258 directly and with the amide nitrogens of Gly 260, Ala 261 and Ile 262 through a water molecule (Fig. 3c). This glutamate is conserved (Fig. 1a) and in GAT1 its replacement by aspartate renders the transporter inactive²⁷.

The aliphatic side chain of leucine, by contrast with the α substituents, is cradled within a hydrophobic pocket created by the side-chain atoms of Val 104 and Tyr 108 in TM3, Phe 253, Ser 256 and Phe 259 in TM6, and Ser 355 and Ile 359 in TM8. The complementarity of the van der Waals surface formed by these residues to the side chain of L-leucine (Fig. 3d) is consistent with the specificity of LeuT_{Aa} for its namesake substrate. Indeed, in a competition experiment 0.1 mM cold L-leucine quantitatively blocks uptake, whereas 0.1 mM of either glycine or tryptophan are much less effective competitors, reducing the uptake of [³H]leucine by only $\sim 30\%$ (Supplementary Fig. 4b).

The location of the substrate binding site in LeuT_{Aa}, together with amino acid sequence information on eukaryotic homologues, allows us to understand why some eukaryotic transporters are specific for glycine and GABA, molecules that possess amino and carboxy groups, whereas other transporters are specific for biogenic amines, including dopamine, serotonin and norepinephrine, molecules that lack a carboxy group. Amino acid sequence analysis illustrates that a key difference between amino acid and biogenic amine transporters is that the former have a glycine and the latter have an aspartate at position 24 in the LeuT_{Aa} sequence¹⁰ (Fig. 1a). Upon modelling an aspartate at glycine 24, we can position the β -carboxy group of the aspartate to within $1\text{ \AA}$ of the carboxy group of the substrate leucine and within $3\text{ \AA}$ of a sodium ion. Therefore, we propose that in the biogenic amine transporters the aspartate's carboxy group supplants the carboxy group of amino acid substrates and coordinates a sodium ion. In addition, the carboxy group of the aspartate in the biogenic amine transporters may form a hydrogen bond to the amine group of the substrate and the hydroxyl group of Tyr 108.

Identification of residues shaping the leucine binding site in LeuT_{Aa} illuminates determinants of substrate specificity in the eukaryotic homologues. In the glycine transporter GlyT1b, for example, residues at equivalent positions to those surrounding the isopropyl moiety of leucine in LeuT_{Aa} are replaced with amino acids of a larger size or different shape, such as Val 104 to Ile, Phe 259 to Trp and Ile 359 to Leu (Fig. 1a). Together, these changes reduce the volume of the binding pocket, thereby rendering it more complementary to glycine. By contrast, residues in SERT are replaced with smaller amino acids, such as Ser 256 to Gly and Ile 359 to Gly (Fig. 1a), to accommodate the larger serotonin molecule. Phe 259, also a phenylalanine in the biogenic amine transporters (Fig. 1a), may participate in π - π stacking interactions with the aromatic rings

of serotonin, dopamine or norepinephrine in their respective transporters.

Sodium binding sites

In the LeuT_{Aa} structure we have identified two sodium ion binding sites, named Na1 and Na2, within $6\text{ \AA}$ of the α -carbon of the bound leucine molecule, located halfway across the membrane bilayer, at the unwound regions of TM1 and TM6 (Fig. 3a). Because distinguishing between a sodium ion and a water molecule, even at $1.65\text{ \AA}$ resolution, is non-trivial, we based our determination on a number of observations. First, when the two peaks were modelled as oxygen atoms of water molecules there was residual density greater than 3σ in electron density maps calculated with $F_o - F_c$ coefficients, after refinement of atomic positions and temperature factors, even though the temperature factors of the oxygen atoms refined to values that were lower (10.41 and $8.81\text{ \AA}^2$) than those of surrounding protein atoms. Second, when the two electron density features were modelled as sodium ions, the $F_o - F_c$ residual density disappeared, and the temperature factors refined to values (17.28 and $15.78\text{ \AA}^2$) comparable to those of neighbouring protein atoms. Third, valence calculations²⁸ yielded a sodium-specific valence (ν_{Na^+}) of 1.32 for Na1 and 1.34 for Na2, in agreement with the values expected for bound sodium ions ($\nu_{\text{Na}^+} \geq 1.0$). By contrast, Li^+ , K^+ and Mg^{2+} yielded valence values far from the expected figures. Fourth, the distances between the sodium ions and coordinating atoms were appropriate for sodium but too short for water or potassium. Valence calculations on all the other water molecules built into the electron density did not suggest the presence of additional cations.

The two sodium ions have key roles in stabilizing the LeuT_{Aa} core, the unwound structures of TM1 and TM6, and the bound leucine molecule. Octahedral coordination of Na1 is provided by the leucine carboxy oxygen, the carbonyl oxygens of Ala 22 (TM1) and Thr 254 (TM6), the side-chain carbonyl oxygens of Asn 27 (TM1) and Asn 286 (TM7), and the hydroxyl oxygen of Thr 254 (TM6; Fig. 4a). In the biogenic amine transporters, an aspartate residue, located at position 24 in the LeuT_{Aa} sequence, probably coordinates a sodium ion equivalent to Na1. Na2 is positioned between the TM1 unwound region and TM8, about $7.0\text{ \AA}$ from Na1 and $5.9\text{ \AA}$ from the α -carbon of bound leucine (Fig. 3a). Trigonal bi-pyramidal coordination to Na2 is achieved by carbonyl oxygens from Gly 20 and Val 23 (TM1), Ala 351 (TM8), and the hydroxyl oxygens from Thr 354 and Ser 355 (TM8; Fig. 4b). Notably, coordination for both Na1 and Na2 is accomplished by partial charges from the protein, with the carboxy group of the bound leucine being the only ligand bearing a formal charge.

What do the sodium binding sites in the LeuT_{Aa} structure tell us about the mechanism by which Na^+/Cl^- -dependent transporters preferentially bind sodium instead of potassium, ions that differ in ionic radius by only $\sim 0.38\text{ \AA}$? We suggest that the primary mechanism is that the sodium binding sites in LeuT_{Aa} are simply too 'small' to accommodate the larger potassium ion. Indeed, the mean sodium ion–ligand distance in LeuT_{Aa} is $2.28\text{ \AA}$ ($\sigma = 0.15\text{ \AA}$), a distance that agrees well with sodium ion–oxygen distances obtained from high-resolution crystal structures²⁹. By comparison, the mean potassium–oxygen distances are $2.84\text{ \AA}$ ²⁹. The ion binding sites in LeuT_{Aa} are devoid of water (Supplementary Fig. 5) and in order for the protein to compensate for the unfavourable dehydration of sodium ions, it must form precise binding sites complementary in size and charge. Because the sodium ions are situated within the core of the protein, the binding sites are constrained from expanding or contracting by a large number of hydrogen-bonding and packing interactions, and are unable to accommodate ions of different size or valence. In fact, the residues surrounding the sodium and leucine binding sites are the most well defined in the entire structure, as indicated by low crystallographic temperature factors (data not shown).

The intimate disposition of the sodium and leucine binding sites implies a mechanism for the coupling of sodium and substrate

transport. We suggest that binding of sodium ions is required to organize the substrate binding site, which is partially formed by the potentially flexible and unwound regions of TM1 and TM6, for high-affinity substrate binding. The coordination of the two sodium ions by residues in TM7 and TM8 further stabilizes the architecture of the substrate-binding site (Fig. 4).

The LeuT_{Aa} structure, together with functional and amino acid sequence data on eukaryotic homologues, suggest that the Na1 site is present in eukaryotic NSS family members. First, TM1, TM6 and TM7, the helices surrounding Na1, are implicated in sodium binding in GAT1 (ref. 30), SERT³¹ and DAT³². Second, coordinating residues are conserved or conservatively substituted in NSS transporters (Fig. 1a), and in SERT the residue equivalent to Asn 286 is involved in sodium binding³¹. In an invertebrate NSS homologue that can use either Na⁺ or K⁺ to drive uptake of neutral amino acids, the corresponding residue is an aspartate, and mutation of this residue alters ion selectivity³³, suggesting that this transporter also shares the Na1 site. Residues coordinating Na2, by contrast, are less well conserved. For example, Thr 354 is a glycine in GlyT1b and an aspartate in GAT1 and the biogenic amine transporters (Fig. 1a). Whereas the aspartate residue probably participates in sodium coordination, in GlyT1b the loss of the threonine hydroxyl leaves only four coordinating atoms and begs the question of whether this site is present in GlyT1b.

Sodium ion to substrate stoichiometry required for transport varies among eukaryotic transporters and is certainly related to the

number of sodium ion binding sites. Although the norepinephrine transporter (NET)¹⁵ and SERT¹⁵ both transport Na⁺ and their respective substrate in a ratio of 1:1 (ref. 15), the sodium-to-substrate ratio in GAT1 (ref. 16) and GlyT1b¹⁷ is 2:1, and in GlyT2a it is 3:1 (ref. 17). We suggest that the Na1 site is a common binding site for the co-transported sodium ion in all NSS members and that those transporters with 2:1 and 3:1 stoichiometries have additional sites that may include the Na2 site. At present, the location of a third sodium site is unknown.

Mechanistic implications

Na⁺/Cl⁻-dependent transporters are membrane-spanning proteins that catalyse the movement of molecules (substrates) and ions across lipid bilayer membranes. These transporters, like many ion channels, have selective binding sites for permeant molecules and ions, but unlike most ion channels, transporters are believed to possess two 'gates' that can alternately allow access to the binding sites from either side of the membrane bilayer, as judged by their function³⁴. The LeuT_{Aa} crystal structure defines the residues and protein domains that comprise the extracellular and intracellular gates, and it suggests conformational changes that might occur during the transport cycle.

The conformation of the LeuT_{Aa} protein reported here is one in which the leucine and sodium sites are occluded from solution on both the extracellular and cytoplasmic sides of the protein; that is, both the extracellular and intracellular gates are closed. Access to the extracellular solution from the leucine and sodium binding sites is primarily obstructed by Tyr 108 and Phe 253, and layered on top of these aromatic residues is a conserved charged pair composed of Arg 30 and Asp 404, which together define at least part of the extracellular gate (Fig. 5a). Arg 30 and Asp 404 interact via a pair of water molecules but could form a salt bridge upon rearrangement of the guanidinium group (Fig. 5b). Arg 30 also forms a cation- π interaction with Phe 253, and is ultimately connected to Na1 via a hydrogen-bond network that includes Gln 250. These residues are highly conserved in the NSS family (Fig. 1a), with residues equivalent to Arg 30 and Asp 404 being important for eukaryotic transporter function^{14,35}. The charged pair lies at the bottom of the extracellular cavity, lined with small pockets and protrusions, defined by TM1b, TM3, TM6a, TM8, TM10 and the 'V-shaped' tip of EL4 (Figs 3a and 5a). We suggest that when the extracellular gate is open, Arg 30 and Asp 404 may move apart, disrupting their interaction. When the gate is closed, Arg 30 and Asp 404 form a water-mediated salt link, stabilizing the extracellular gate in a closed conformation.

In contrast to the extracellular gate, where just a few residues obstruct the leucine and sodium ion binding sites, the cytoplasmic gate is far more substantial, being composed of ~20 Å of ordered protein structure (Fig. 5a). Directly beneath the leucine and sodium binding sites are TM1a, TM6b and TM8, with these main-chain scaffolds themselves blocking cytoplasmic access to the binding sites. Near the cytoplasmic face of the protein, another conserved charged pair, Arg 5 and Asp 369, forms an important and partially buried salt bridge. Arg 5 also forms a hydrogen bond with Ser 267 and a cation- π interaction with Tyr 268 (Fig. 5c). Adjacent to these residues lies Trp 8, the indole ring of which fits snugly in a hydrophobic pocket created by TM1a and TM6b, with its Ne nitrogen atom hydrogen bonding to the carbonyl oxygen of Tyr 265 in TM6b (Fig. 5c). The conserved Trp 8 stabilizes the conformation of TM1a-TM6b and anchors the amino terminus, including Arg 5. All these residues are strictly conserved among other NSS family members (Fig. 1a), and mutation of residues equivalent to either Arg 5 or Trp 8 in GAT1 (ref. 36) and Asp 369 (ref. 37) or Tyr 268 (ref. 38) in DAT impairs transport.

How might the extracellular and cytoplasmic gates open? The architecture of LeuT_{Aa} provides tantalizing clues to possible movements of the protein backbone. If we consider the unwound regions as 'joints', we suggest that the extracellular and cytoplasmic helical segments, that is TM1b-TM6a and TM1a-TM6b, respectively, might

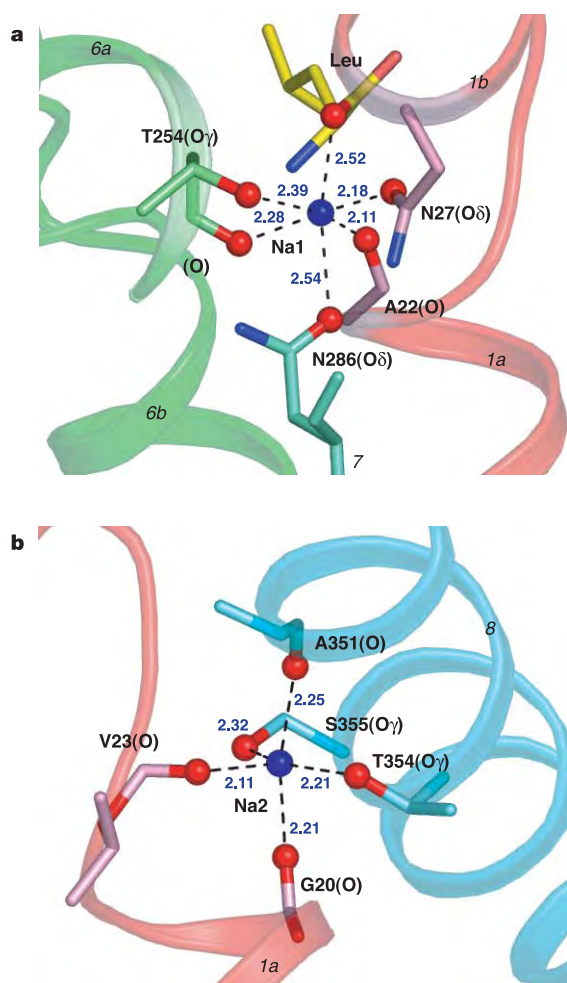


Figure 4 | Sodium ion binding sites. **a**, Na1 and residues from TM1a, TM1b, TM6a and TM7 together with bound leucine are shown. **b**, Na2 and residues from TM1a and TM8 are shown. Distances (Å) are in blue letters.

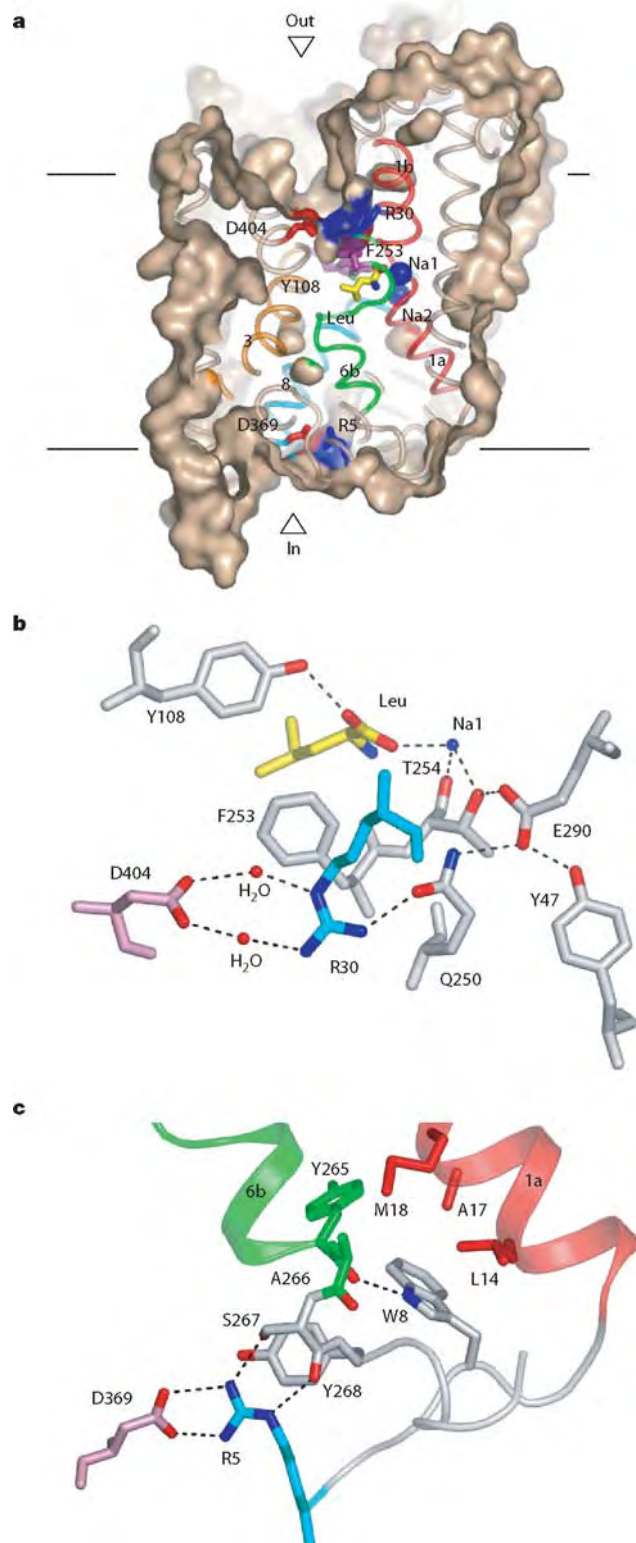


Figure 5 | Extracellular and cytoplasmic gates. **a**, Slice through the surface of LeuT_{Aa}, viewed parallel to the membrane, showing the extracellular cavity. Connolly surface of LeuT_{Aa} is shown in beige. L-Leucine, Tyr 108, Phe 253 and the two charged pairs (Arg 30–Asp 404 and Arg 5–Asp 369) are depicted as stick models in yellow for leucine, purple for aromatic residues, blue for arginines and red for aspartates. **b**, **c**, Key interacting residues at the extracellular (**b**) and at the cytoplasmic (**c**) gate.

move relative to TM3 and TM8, helping to open and close the outward and inward gates (Fig. 6). Binding and unbinding of substrate and ions to the unwound joints may stabilize the helices in different conformational states. In calcium ATPase, for example, binding and release of calcium ions to unwound regions in transmembrane helices effects structural rearrangements³⁹. Movement in TM1 may be accompanied by a conformational change in the linker between TM1 and TM2, which contains the highly conserved motif NGGGAF. In GLYT2a⁴⁰ and SERT⁴¹, accessibility studies have revealed that this region moves during the transport cycle. We suggest that TM3 and TM8 may also shift their positions or rotate slightly. In addition, the extracellular and cytoplasmic entries could expand upon movement of surface-exposed elements, including the N terminus, IL1, EL2 and EL4. Indeed, construction of chimaeras between NET and SERT revealed that both EL2 (ref. 42) and EL4 (ref. 43) participate in conformational changes during transport. Finally, the surrounding transmembrane segments, TM2, TM7, TM10 and TM11, may accommodate movement of TM1, TM3, TM6 and TM8 because they have interruptions in the middle of their helical structures caused by Gly 55 and Pro 57 in TM2, Gly 294 in TM7, Gly 408 in TM10 and Pro 457 in TM11, most of which are conserved residues in the NSS family (Fig. 1a).

Discussion

The structure of LeuT_{Aa}, together with that of the prokaryotic glutamate transporter homologue Glt_{Ph} of *Pyrococcus horikoshii*⁴⁴, a representative of the other family of plasma membrane neurotransmitter transporters, reveals important common principles between them, although their overall folds are markedly different. First, both structures have captured an occluded state, therefore suggesting that mechanisms of transport for these transporters must include at least three iconic states: open to outside, occluded, and open to inside. Second, key portions of the substrate and ion binding sites are composed of extended elements of polypeptide insinuated in the middle of disrupted transmembrane α -helices, thereby allowing main-chain atoms to participate in hydrogen bonding interactions. Finally, substrate and sodium binding sites (LeuT_{Aa}) or residues implicated in sodium binding (Glt_{Ph}) are located close in three-dimensional space, thereby demonstrating that the long-described thermodynamic coupling of substrate and ion transport in these two families of sodium-dependent transporters is the consequence of direct or nearly direct interactions.

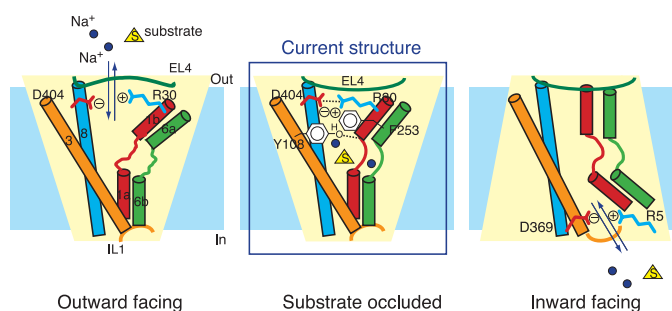


Figure 6 | Speculative transport mechanism. Schematic drawing of a possible conformational change upon substrate/sodium ion transport. The left panel shows the outward-facing state. TM1a and TM6b assume the closed arrangement, whereas TM1b and TM6a adopt the open one. The middle panel shows the substrate-occluded state, which corresponds to the current crystal structure. TM1a and TM6b assume the closed arrangement, whereas TM1b and TM6a adopt a partially open one with some residues blocking the permeation pathway. The right panel shows the inward-facing state. TM1b and TM6a assume the closed arrangement, whereas TM1a and TM6b adopt the closed one, to open the pathway to the cytoplasm.

METHODS

Crystallization. The DNA coding for *A. aeolicus* LeuT_{AA} was amplified by polymerase chain reaction from genomic DNA and cloned into a pET16b derivative that included a C-terminal octahistidine tag and a thrombin site. The protein was produced in *Escherichia coli* C41 cells cultured in Terrific broth after induction at an absorbance at 600 nm (A_{600}) of 0.6 with 0.1 mM isopropyl- β -D-thiogalactopyranoside for 20 h at 20 °C. Cell membranes were isolated from disrupted cells and solubilized with 40 mM dodecylmaltoside (DDM). LeuT_{AA} was eluted from a metal ion affinity chromatography column in buffer composed of 20 mM Tris-HCl (pH 8.0), 190 mM NaCl, 10 mM KCl, 300 mM imidazole and 1 mM DDM. After thrombin digestion, the protein was loaded onto a size exclusion column equilibrated in 20 mM Tris-HCl (pH 8.0), 190 mM NaCl, 10 mM KCl, and 40 mM *n*-octyl- β -D-glucopyranoside (β -OG). Purified protein was concentrated and dialysed overnight at 4 °C against 10 mM Tris-HCl (pH 8.0), 45 mM NaCl, 5 mM KCl and 40 mM β -OG before crystallization. Selenomethionine-labelled protein (SeMet-LeuT_{AA}) was expressed in C41 cells as previously reported⁴⁵, using 1 mM β -mercaptoethanol in all buffers to minimize oxidation.

LeuT_{AA} crystals were grown by hanging-drop vapour diffusion at 18 °C by mixing protein solution (~ 7 mg ml⁻¹) with a reservoir solution composed of 0.1 M HEPES-NaOH (pH 7.0), 18–22% PEG 550 MME, and 0.2 M NaCl. Before freezing in liquid nitrogen, crystals were cryoprotected with a reservoir solution containing 35% PEG 550 MME and 40 mM β -OG. The crystals belong to the space group of C2 with cell dimensions of $a = 87.9$ Å, $b = 86.3$ Å, $c = 81.0$ Å and $\beta = 95.7^\circ$.

Structure determination. X-ray diffraction data sets were collected using an ADSC QUANTUM 315 detector at the Advanced Light Source 8.2.2 and processed with HKL2000. Phases were obtained by the multiwavelength anomalous dispersion method¹⁹ using a three-wavelength data set from a SeMet-LeuT_{AA} crystal. The program SOLVE⁴⁶ was used to determine selenium sites and initial phases, which were improved by solvent flattening and histogram matching using DM⁴⁷. The initial structure was automatically built by Arp/wArp⁴⁸, subsequent manual rebuilding was accomplished with TURBO-FRODO, and the model was refined with CNS⁴⁹.

Transport assay. Proteoliposomes were reconstituted as described previously⁵⁰ using LeuT_{AA} as purified for crystallization except with 1 mM DDM as the detergent. L-Leucine uptake was initiated by diluting liposomes loaded with buffer 1 (20 mM HEPES-Tris, pH 7.4, 100 mM potassium gluconate) supplemented with 1 μ M valinomycin into a 200-fold excess of buffer 2 (20 mM HEPES-Tris, pH 7.4, 0.1 μ M L-[3,4,5-³H(N)]leucine) also containing 1 μ M valinomycin and either 100 mM NaCl, 100 mM sodium gluconate, or 100 mM KCl at 20 °C. Reactions were quenched by diluting each assay tenfold with ice-cold buffer 1, followed by filtration through nitrocellulose filters. The filters were washed twice with 1 ml of ice-cold buffer 1, placed in scintillation fluid, and counted on a scintillation counter. Background counts were subtracted from each data point and were estimated by averaging [³H]leucine counts over a 10-min time window using liposomes prepared in the absence of LeuT_{AA} and reaction conditions as described above. Specific transport activity was estimated from the protein concentration before reconstitution, assuming an incorporation efficiency of 100%. The experiments probing valinomycin dependence and substrate competition are described in Supplementary Information.

Received 23 May; accepted 4 July 2005.

Published online 24 July 2005.

- Vanhatalo, S. & Soinila, S. The concept of chemical neurotransmission – variations on the theme. *Ann. Med.* **19**, 151–158 (1998).
- Masson, J., Sagne, C., Hamon, M. & Mestikawy, S. E. Neurotransmitter transporters in the central nervous system. *Pharm. Rev.* **51**, 439–464 (1999).
- Hahn, M. K. & Blakely, R. D. Monoamine transporter gene structure and polymorphisms in relation to psychiatric and other complex disorders. *Pharmacogenomics J.* **2**, 217–235 (2002).
- Richerson, G. B. & Wu, Y. Role of the GABA transporter in epilepsy. *Adv. Exp. Med. Biol.* **548**, 76–91 (2004).
- Amara, S. G. & Sonders, M. S. Neurotransmitter transporters as molecular targets for addictive drugs. *Drug Alcohol Depend.* **51**, 87–96 (1998).
- Krogsgaard-Larsen, P., Frolund, B. & Frydenvang, K. GABA uptake inhibitors. Design, molecular pharmacology and therapeutic aspects. *Curr. Pharm. Des.* **6**, 1193–1209 (2000).
- Barker, E. L. & Blakely, R. D. In *Psychopharmacology—the Fourth Generation of Progress* (eds Bloom, F. E. & Kupfer, D. J.) (Raven Press, New York, 2000).
- Guastella, J. et al. Cloning and expression of a rat brain GABA transporter. *Science* **249**, 1303–1306 (1990).
- Nelson, N. The family of Na⁺/Cl⁻-dependent neurotransmitter transporters. *J. Neurochem.* **71**, 1785–1803 (1998).
- Torres, G. E., Gainetdinov, R. R. & Caron, M. G. Plasma membrane monoamine transporters: structure, regulation, and function. *Nature Rev. Neurosci.* **4**, 13–25 (2003).
- Chen, J. G., Liu-Chen, S. & Rudnick, G. Determination of external loop topology in the serotonin transporter by site-directed chemical labeling. *J. Biol. Chem.* **273**, 12675–12681 (1998).
- Bismuth, Y., Kavanaugh, M. P. & Kanner, B. I. Tyrosine 140 of the γ -aminobutyric acid transporter GAT-1 plays a critical role in neurotransmitter recognition. *J. Biol. Chem.* **272**, 16096–16102 (1997).
- Chen, J. G., Sachpatzidis, A. & Rudnick, G. The third transmembrane domain of the serotonin transporter contains residues associated with substrate and cocaine binding. *J. Biol. Chem.* **272**, 28321–28327 (1997).
- Cao, Y., Li, M., Mager, S. & Lester, H. A. Amino acid residues that control pH modulation of transport-associated current in mammalian serotonin transporters. *J. Neurosci.* **18**, 7739–7749 (1998).
- Rudnick, G. in *Neurotransmitter Transporters: Structure, Function, and Regulation* (ed. Reith, E. A.) 25–52 (Humana Press, Totowa, New Jersey, 2002).
- Kavanaugh, M. P., Arriza, J. L., North, R. A. & Amara, S. G. Electrogenic uptake of γ -aminobutyric acid by a cloned transporter expressed in *Xenopus* oocytes. *J. Biol. Chem.* **267**, 22007–22009 (1992).
- Roux, M. & Supplisson, S. Neuronal and glial glycine transporters have different stoichiometries. *Neuron* **25**, 373–383 (2000).
- Androutsellis-Theotokis, A. et al. Characterization of a functional bacterial homologue of sodium-dependent neurotransmitter transporters. *J. Biol. Chem.* **278**, 12703–12709 (2003).
- Hendrickson, W. A. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
- Murata, K. et al. Structural determinants of water permeation through aquaporin-1. *Nature* **407**, 599–605 (2000).
- Hirai, T. et al. Three-dimensional structure of a bacterial oxalate transporter. *Nature Struct. Biol.* **9**, 597–600 (2002).
- Chen, J. G., Liu-Chen, S. & Rudnick, G. External cysteine residues in the serotonin transporter. *Biochemistry* **36**, 1479–1486 (1997).
- Wang, J. B., Moriwaki, A. & Uhl, G. R. Dopamine transporter cysteine mutants: second extracellular loop cysteines are required for transporter expression. *J. Neurochem.* **64**, 1416–1419 (1995).
- Just, H., Sitte, H. H., Schmid, J. A., Freissmuth, M. & Kudlacek, O. Identification of an additional interaction domain in transmembrane domains 11 and 12 that supports oligomer formation in the human serotonin transporter. *J. Biol. Chem.* **279**, 6650–6657 (2004).
- Sitte, H. H., Farhan, H. & Javitch, J. A. Sodium-dependent neurotransmitter transporters: oligomerization as a determinant of transporter function and trafficking. *Mol. Interv.* **4**, 38–47 (2004).
- Ponce, J., Biton, B., Benavides, J., Avenet, P. & Aragon, C. Transmembrane domain III plays an important role in ion binding and permeation in the glycine transporter GLYT2. *J. Biol. Chem.* **275**, 13856–13862 (2000).
- Keshet, G. I. et al. Glutamate-101 is critical for the function of the sodium and chloride-coupled GABA transporter GAT-1. *FEBS Lett.* **371**, 39–42 (1995).
- Nayal, M. & Di Cera, E. Valence screening of water in protein crystals reveals potential Na⁺ binding sites. *J. Mol. Biol.* **256**, 228–234 (1996).
- Harding, M. M. Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. D* **58**, 872–874 (2002).
- Mager, S. et al. Ion binding and permeation at the GABA transporter GAT1. *J. Neurosci.* **16**, 5405–5414 (1996).
- Penado, K. M., Rudnick, G. & Stephan, M. M. Critical amino acid residues in transmembrane span 7 of the serotonin transporter identified by random mutagenesis. *J. Biol. Chem.* **273**, 28098–28106 (1998).
- Chen, N. & Reith, M. E. Na⁺ and the substrate permeation pathway in dopamine transporters. *Eur. J. Pharmacol.* **479**, 213–221 (2003).
- Mari, S. A. et al. Aspartate 338 contributes to the cationic specificity and to driver-amino acid coupling in the insect cotransporter KAT1. *Cell. Mol. Life Sci.* **61**, 243–256 (2004).
- Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
- Pantanowitz, S., Bendahan, A. & Kanner, B. I. Only one of the charged amino acids located in the transmembrane α -helices of the γ -aminobutyric acid transporter (subtype A) is essential for its activity. *J. Biol. Chem.* **268**, 3222–3225 (1993).
- Bennett, E. R., Su, H. & Kanner, B. I. Mutation of arginine 44 of GAT-1, a (Na⁺ + Cl⁻)-coupled γ -aminobutyric acid transporter from rat brain, impairs net flux but not exchange. *J. Biol. Chem.* **275**, 34106–34113 (2000).
- Loland, C. J., Granas, C., Javitch, J. A. & Gether, U. Identification of intracellular residues in the dopamine transporter critical for regulation of transporter conformation and cocaine binding. *J. Biol. Chem.* **279**, 3228–3238 (2004).
- Loland, C. J., Norregaard, L., Litman, T. & Gether, U. Generation of an activating Zn²⁺ switch in the dopamine transporter: mutation of an intracellular tyrosine constitutively alters the conformational equilibrium of the transport cycle. *Proc. Natl Acad. Sci. USA* **99**, 1683–1688 (2002).
- Toyoshima, C. & Nomura, H. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature* **418**, 605–611 (2002).
- Lopez-Corcuera, B., Nunez, E., Martinez-Maza, R., Geerlings, A. & Aragon, C. Substrate-induced conformational changes of extracellular loop 1 in the glycine transporter GLYT2. *J. Biol. Chem.* **276**, 43463–43470 (2001).

41. Sato, Y., Zhang, Y. W., Androutsellis-Theotokis, A. & Rudnick, G. Analysis of transmembrane domain 2 of rat serotonin transporter by cysteine scanning mutagenesis. *J. Biol. Chem.* **279**, 22926–22933 (2004).
42. Stephan, M. M., Chen, M. A., Penado, K. M. & Rudnick, G. An extracellular loop region of the serotonin transporter may be involved in the translocation mechanism. *Biochemistry* **36**, 1322–1328 (1997).
43. Smicun, Y., Campbell, S. D., Chen, M. A., Gu, H. & Rudnick, G. The role of external loop regions in serotonin transport. *J. Biol. Chem.* **274**, 36058–36064 (1999).
44. Yernool, D., Boudker, O., Jin, Y. & Gouaux, E. Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* **431**, 811–818 (2004).
45. Guerrero, S. A., Hecht, H. J., Hofmann, B., Biebl, H. & Singh, M. Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Appl. Microbiol. Biotechnol.* **56**, 718–723 (2001).
46. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
47. Collaborative Computational Project, No. 4, The CCP4 suite: program for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
48. Perrakis, A., Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463 (1999).
49. Brünger, A. T. et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
50. Yernool, D., Boudker, O., Foltá-Stogniew, E. & Gouaux, E. Trimeric subunit stoichiometry of the glutamate transporters from *Bacillus caldotenax* and *Bacillus stearothermophilus*. *Biochemistry* **42**, 12981–12988 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We appreciate the beamtime and the assistance of the personnel at beamline 8.2.2 of the Advanced Light Source and beamlines X4A, X12B and X29 of the National Synchrotron Light Source; B. Honig and L. Forrest for help with sequence alignment; M. A. Gawinowicz for mass spectrometry and free amino acid analysis; J. Moon for help with bacterial culture and crystallization; H. Furukawa for assistance with vector construction and light scattering experiments; O. Boudker for liposome reconstitution; and D. Yernool, O. Boudker and R. Ryan for comments on the manuscript. A.Y. is on leave from the Laboratory for Structural Biochemistry, RIKEN Harima Institute at SPring-8, Japan. S.K.S. is supported by an NIH NRSA postdoctoral fellowship. The work was supported by the NIH. E.G. is an investigator with the Howard Hughes Medical Institute.

Author Information The coordinates for the structure have been deposited in the Protein Data Bank under the accession code 2A65. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.G. (jeg52@columbia.edu).

LETTERS

Differentiation of the asteroid Ceres as revealed by its shape

P. C. Thomas¹, J. Wm. Parker², L. A. McFadden³, C. T. Russell⁴, S. A. Stern², M. V. Sykes⁵ & E. F. Young²

The accretion of bodies in the asteroid belt was halted nearly 4.6 billion years ago by the gravitational influence of the newly formed giant planet Jupiter. The asteroid belt therefore preserves a record of both this earliest epoch of Solar System formation and variation of conditions within the solar nebula. Spectral features in reflected sunlight indicate that some asteroids have experienced sufficient thermal evolution to differentiate into layered structures¹. The second most massive asteroid—4 Vesta—has differentiated to a crust, mantle and core^{2,3}. 1 Ceres, the largest and most massive asteroid, has in contrast been presumed to be homogeneous, in part because of its low density, low albedo and relatively featureless visible reflectance spectrum, similar to carbonaceous meteorites that have suffered minimal thermal processing⁴. Here we show that Ceres has a shape and smoothness indicative of a gravitationally relaxed object. Its shape is significantly less flattened than that expected for a homogeneous object, but is consistent with a central mass concentration indicative of differentiation. Possible interior configurations include water-ice-rich mantles over a rocky core.

We obtained Hubble Space Telescope (HST) observations during nine orbits in December 2003 and January 2004 using the High Resolution Channel (HRC) of the Advanced Camera for Surveys (ACS)⁵. Six consecutive HST orbits were used to provide full coverage of Ceres' 9.075-h rotation period (Fig. 1), and three additional orbits separated by 120° in rotational phase were used to obtain higher spatial resolution by making small offsets (dithering) of the telescope between exposures. The data were processed through the standard HST pipeline, followed by processing with software using standard HRC functions to remove geometric distortion.

Limb profiles were measured for 217 images using techniques that locate sharp contrast boundaries to ~0.1 pixel (ref. 6). We find that the shape of Ceres is rotationally symmetric to the limit of measurement accuracy (Fig. 2a, b), and is well described by an oblate spheroid of axes $a = 487.3 \pm 1.8$ km and $b = 454.7 \pm 1.6$ km (1σ). Individual limb profiles are well fitted by ellipses: combining data for all 217 limbs, the root mean square radial error is 0.16 pixels, or ~4.8 km (Fig. 2c).

Because the equatorial axes are the same within <2 km, and mean residuals to ellipses fitted to the limbs are <5 km, we conclude that Ceres is a relaxed object; that is, its shape is determined by hydrostatic equilibrium. A globally relaxed object can still have topography, such as craters, because relaxation timescales are inversely proportional to the wavelength of the topography⁷. The limb topography from our data are smoothed over the 30-km pixel size, and we would probably not detect topography <60 km across. Occultation data⁸, which effectively sample at points, have average residuals of about 5 km, with maximum topography about 10 km. Such relief is consistent with our data.

If water ice is important in the rheology of Ceres' near-surface region, relief in large impact basins on icy satellites is our best practical guide to how much relief there should be on Ceres. Because larger craters are relatively more shallow than small ones, even craters over 200 km across have depths of less than 10 km (ref. 9). As an example, there are excellent limb profile data on the 400-km crater Odysseus on Saturn's moon Tethys (mean radius of 535 km) that show maximum relief of about 8 km (ref. 9); smoothed to 30 km spacing, this would show ~5 km relief. Such relief on Ceres is difficult to detect by current means, and might be reduced by the slightly higher expected temperatures compared to Tethys. Thus, if ice is involved in the rheology of Ceres, it could retain a substantial crater record without affecting the basic relaxed shape. For comparison, Vesta, which is clearly a differentiated object with a basaltic crust, does not fit an equilibrium form^{10,11}, either in the match of overall dimensions to an equilibrium form (10 km variations), or in the limb topography (up to 20 km of relief¹⁰). Given the similar gravity on the two objects (~25 cm s⁻²), the difference in shapes probably reflects

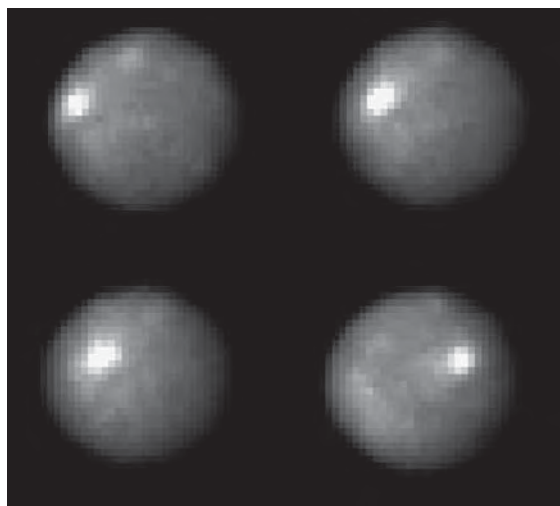


Figure 1 | Rotation of a bright spot on Ceres. Advanced Camera for Surveys (ACS) images of Ceres, showing rotation of a bright spot. The images are strongly contrast-enhanced to emphasize the feature; actual contrasts on the surface are only a few percent. The illuminated limb is the left side, centred 17° up from horizontal. The limb ellipse fits, combined with tracking of the bright spot visible in 51 images covering 93° of rotation, allow us to find a spin vector of right ascension (RA) 291° and declination (dec.) 59°. The uncertainty of ~5° clarifies ambiguities among previous measurements^{28,29}. This spin pole means that Ceres has an obliquity of only ~3°, which is consistent with previous interpretation, but reduces the uncertainty.

¹Center for Radiophysics and Space Research, Cornell University, Ithaca, New York 14853, USA. ²Department of Space Studies, SwRI, 1050 Walnut Street, Boulder, Colorado 80302, USA. ³Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA. ⁴IGPP & ESS, University of California Los Angeles, Los Angeles, California 90095, USA. ⁵Planetary Science Institute, 1700 East Fort Lowell, Tucson, Arizona 85719-2395, USA.

different rheologies, with Ceres unable to sustain as much topography as Vesta. Smaller asteroids for which there are good shape data are irregular in form, and expectations of equilibrium forms among many asteroids have not been met¹². Thus, Ceres is as yet the only known relaxed asteroid.

The difference between the long and short axes of an equilibrium spheroid is directly proportional to the square of the rotation rate, is inversely proportional to the mean density^{13,14}, and is affected by the distribution of mass. Thus, the shape may be used to investigate global internal properties. Based on the maximum and minimum values of recent measurements of Ceres' mass of $(9.47 \pm 0.05) \times 10^{20}$ kg and $(9.35 \pm 0.08) \times 10^{20}$ kg (refs 11, 15), we use an average mass of $(9.395 \pm 0.125) \times 10^{20}$ kg. With our mean radius of 476.2 ± 1.7 km (1σ , calculated from each observed long and short axis), this mass corresponds to a density of $2,077 \pm 36$ kg m⁻³. This density falls within the range of bulk densities for Murchison-like (CM) carbonaceous chondrites¹⁶, but is larger than the density of $1,300$ kg m⁻³ of the C-type asteroid Mathilde¹⁷, whose low density is attributed to interior voids. Porosity is not expected to be globally important for Ceres, as the central pressure of a body such as Mathilde is reached within ~ 300 m of the surface of Ceres.

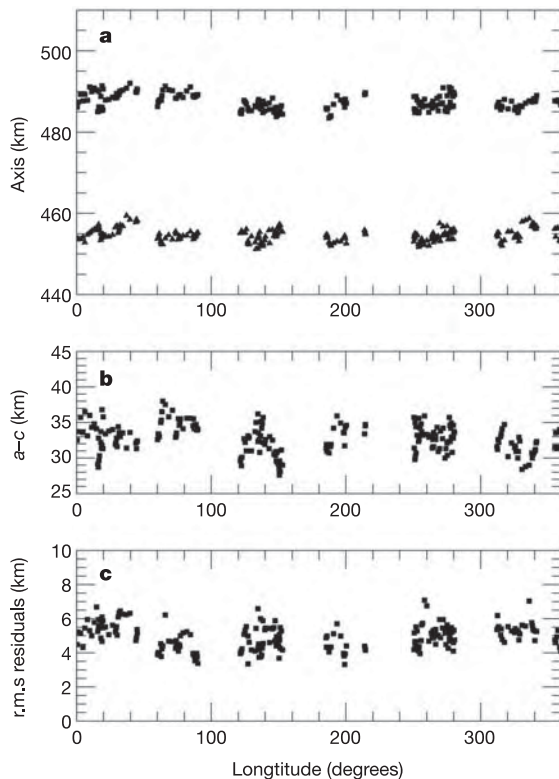


Figure 2 | The shape of Ceres. **a**, Short (polar) and long (equatorial) projected axes of Ceres plotted versus an arbitrary reference viewing longitude. The pixel scale of 0.025 arcsec in the geometrically corrected images is about 30 km per pixel at the geocentric distance of Ceres (~ 1.65 AU) during the observations. The low solar phase angles of 5.4° – 7.5° allow use of both the limb and terminator positions, but induce radius errors between the terminator and ideal body outline. Terminator portions of raw profiles were corrected for predicted position dependent upon solar phase and azimuth in the image. These corrections result in changes to fit radii of 1–2 km, which are ~ 0.03 – 0.06 pixels. The sub-observer latitudes on Ceres were -0.4° to $+2.1^\circ$ N. This equatorial viewing geometry means the projected axes are essentially identical to the actual shape. Three filter bands were used, centred at 535, 335 and 223 nm (ref. 30); averaged by filter, the fit axes differ by less than 1 km. **b**, Difference between long (a) and short (c) axes of Ceres as a function of arbitrary observer longitude. **c**, Smoothness of limb of Ceres. Root mean square residuals of fits to ellipses are shown as a function of observer longitude.

For a mean density of $2,077$ kg m⁻³, the equilibrium shape ($a - c$) would be 39.7 km (ref. 14). Our value of $a - c$ of 32.6 ± 1.95 km (1σ) differs from this value by more than 3σ . Thus, the observed flattening is inconsistent with that expected for a homogeneous body¹⁴. The same test was previously applied to the occultation data⁸ but used a higher mass, and thus obtained a result consistent with a homogeneous Ceres. However, we must evaluate the possible systematic errors to determine how relevant the formal uncertainties are.

The primary contributor of error in the shape is any difference in systematic errors in detecting the limb along the long and short axes. Consistent errors in the same sense will not affect the shape solution, only the mean radius. We estimate possible systematic errors in the measurement of the long and short axes by comparison to past experience⁶ with the limb-finding routine on sections of limbs that include greatly different albedos and different scattering functions. Large photometric variations cause systematic changes of individual limb coordinates of almost always less than 0.3 pixels. If entirely concentrated on one axis, this error would change the fitted semi-axis by 0.15 pixels. Given that the photometric variation around the disk of Ceres is small, and is not restricted to one projected axis, an assumption of maximum possible errors in $a - c$ of 0.15 pixels, or 4.5 km, is appropriately conservative, and can be regarded as a 3σ error.

The best check of the absolute dimensions of Ceres comes from stellar occultation measurements⁸. This work gave axes of 479.6 ± 2.4 by 454.3 ± 4.5 km if minimizing radial residuals, and 481.6 ± 2.4 by 450.1 ± 2.0 km (1σ values) if minimizing the timing errors. The two solutions give $a - c$ values, 25.3 and 31.5 km, that are less than our value of 32.6 km. The occultation observed only one cross-section of Ceres, but because the HST data show that the long axis varies by only a few kilometres around the body, the comparison

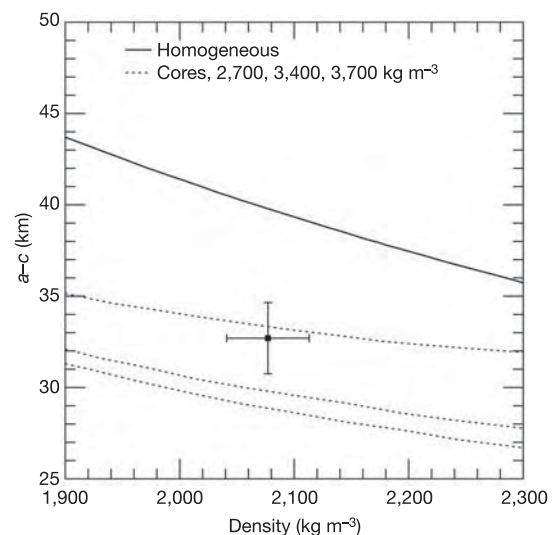


Figure 3 | Interior models of Ceres. Shape and density of Ceres compared to Maclaurin spheroids. Shape is given as the difference between the equatorial, a , and polar, c , radii. Solid curve is the expected shape as a function of density for a relaxed homogeneous body with a rotation period of 9.075 h. Plotted point is the average density of Ceres and the observed shape. Uncertainty in the shape value is the standard deviation (1σ) of all measures of $a - c$. Occultation data⁸ suggest that $a - c$ may be less than our nominal value. Uncertainty in the density is the root sum square of the fractional mass and volume uncertainties. Dashed lines give the equilibrium shape for a simple core/mantle model of rock and water ice: core densities are 2,700, 3,400 and $3,700$ kg m⁻³, going from upper to lower dashed curves. For a homogeneous object the size of Ceres, the numerical determination of shape $a - c$ is within 150 m of the theoretical value¹⁴, indicating the level to which the model predicted shapes are accurate.

is meaningful and shows a mean radius within 6 km of our value.

An $a - c$ value smaller than that expected of a homogeneous object indicates mass concentration towards the centre¹³. We would expect a body of the size and solar distance of Ceres to be composed principally of rock and water ice⁴. However, the mineralogy and density of a non-ice component is not well constrained, and may itself depend upon thermal history^{2,4}. To construct a simple interior model, we assume a range of core densities: one appropriate for the grain density of primitive carbonaceous materials ($2,700 \text{ kg m}^{-3}$), and one that spans recent density measurements of Vesta^{11,15}, $3,400$ to $3,700 \text{ kg m}^{-3}$. For a particular mean density, and assuming a water-ice mantle, a choice of the core density establishes the relative core volume, and the equilibrium shape is determined by numerical methods¹⁸ where gravitational potential is equal over the surface. A rocky core and a water-ice mantle give a shape consistent with our observations (Fig. 3). Assuming the densest core materials and the nominal mean density of $2,077 \text{ kg m}^{-3}$, ice mantles are 110–124 km thick and constitute 24–26% of the body mass. The lighter core material requires a mantle thickness of 66 km and a 16% ice mass. Even the minimum mantle thickness is greater than the likely excavation depths of craters a few hundred kilometres across¹⁹.

An alternative interpretation of the shape would be that Ceres is homogeneous and that its shape stabilized at a longer spin period, $\sim 9.93 \text{ h}$, and has not relaxed after subsequent spin-up to its present period of 9.075 h . Tidal interactions with other asteroids are incapable of this spin-up, and impacts large enough to impart the necessary angular momentum would leave obvious topography²⁰, given an assumption that the shape is 'frozen'.

A water fraction of $\sim 25\%$ is reasonable for objects at this solar distance^{21,22}. However, predictions of how much water ends up in a frozen ice mantle, an ocean, or hydrated silicate minerals are very model dependent³. In any case, the surface of Ceres is dark and devoid of spectral signatures of water ice, though bound water and ammoniated silicates may be present^{23–25}. Near-infrared observations of Ceres indicate evidence of some aqueous alteration, at a temperature not exceeding 400°C (ref. 26). The relaxed state, differentiated structure, and the mean density strongly suggest water ice as the primary mantle constituent. The difference between Ceres' surface and those of icy satellites of Jupiter and Saturn is not surprising, given the much higher heating available at Ceres' distance from the Sun, which makes water ice unstable at the surface²⁷. If water ice has been at the surface of Ceres, it may currently be hidden just below a thin residual layer of clay and dark carbonaceous materials. Our findings indicate that the Dawn mission, which will orbit Ceres in 2015, will find a globally relaxed and differentiated object, but which should retain a visible cratering record, and possibly tectonic features similar to those on some icy satellites.

Received 22 April; accepted 10 June 2005.

1. McSween, H. Y., Ghosh, A., Grimm, E., Wilson, L. & Young, E. D. in *Asteroids III* (eds Bottke, W. et al.) 559–571 (Univ. Arizona Press, Tucson, 2002).
2. McCord, T. B., Adams, J. B. & Johnson, T. V. Asteroid Vesta: Spectral reflectivity and compositional implications. *Science* **168**, 1445–1447 (1970).
3. Ghosh, A. & McSween, H. Y. Jr A thermal model for the differentiation of asteroid 4 Vesta, based on radiogenic heating. *Icarus* **134**, 187–206 (1998).
4. McCord, T. B. & Sotin, C. Ceres: Evolution and current state. *J. Geophys. Res.* **110**, E05009, doi:10.1029/2004JE002244 (2005).
5. Parker, J. Wm. et al. Ceres: High-resolution imaging with HST and the

determination of physical properties. *Adv. Space Res.* (in the press).

6. Thomas, P. C. et al. The shape of Io from Galileo limb measurements. *Icarus* **135**, 175–180 (1998).
7. Johnson, T. V. & McGetchin, T. R. Topography on satellite surfaces and the shape of asteroids. *Icarus* **18**, 612–620 (1973).
8. Millis, R. L. et al. The size, shape, density, and albedo of Ceres from its occultation of BD + 8°471. *Icarus* **72**, 507–518 (1987).
9. Schenk, P. M. Crater formation and modification on the icy satellites of Uranus and Saturn—Depth/diameter and central peak occurrence. *J. Geophys. Res.* **94**, 3813–3832 (1989).
10. Thomas, P. C. et al. Impact excavation on asteroid 4 Vesta: Hubble Space Telescope results. *Science* **277**, 1492–1495 (1997).
11. Viateau, B. & Rapport, N. Mass and density of asteroids (4) Vesta and (11) Parthenope. *Astron. Astrophys.* **370**, 602–609 (2001).
12. Hestroffer, D. On equilibrium shapes among binary asteroids. *Bull. Am. Astron. Soc.* **36**, 861 (2004).
13. Dermott, S. F. Shapes and gravitational moments of satellites and asteroids. *Icarus* **37**, 575–586 (1979).
14. Chandrasekhar, S. *Ellipsoidal Figures of Equilibrium* (Yale Univ. Press, New Haven, 1969).
15. Michalak, G. Determination of asteroid masses—I. (1) Ceres, (2) Pallas and (4) Vesta. *Astron. Astrophys.* **360**, 363–374 (2000).
16. Britt, D. T. & Consolmagno, G. J. Stony meteorite porosities and densities: A review of the data through 2001. *Meteorit. Planet. Sci.* **38**, 1161–1180 (2003).
17. Veverka, J. et al. NEAR's flyby of 253 Mathilde: Images of a C asteroid. *Science* **278**, 2109–2112 (1997).
18. Thomas, P. C. Gravity, tides, and topography on small satellites and asteroids—Application to surface features of the Martian satellites. *Icarus* **105**, 326–344 (1993).
19. Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Univ. Press, New York, 1989).
20. Farinella, P., Davis, D. R., Paolicchi, P., Cellino, A. & Zappala, Z. Asteroid collisional evolution: An integrated model for the evolution of asteroid rotation rates. *Astron. Astrophys.* **253**, 604–614 (1992).
21. Wilson, L., Keil, K., Browning, L. B., Krot, A. N. & Bourcher, W. Early aqueous alteration, explosive disruption, and re-processing of asteroids. *Meteorit. Planet. Sci.* **34**, 541–557 (1999).
22. Grimm, R. E. & McSween, H. Y. Water and the thermal evolution of carbonaceous chondrite parent bodies. *Icarus* **82**, 244–280 (1989).
23. Lebofsky, L. A. et al. The 1.7- to 4.2-micron spectrum of asteroid 1 Ceres—Evidence for structural water in clay minerals. *Icarus* **48**, 453–459 (1981).
24. Feierberg, M. A., Lebofsky, L. A. & Larson, H. P. Spectroscopic evidence for aqueous alteration products on the surfaces of low-albedo asteroids. *Geochim. Cosmochim. Acta* **45**, 971–981 (1981).
25. King, T. V. et al. Evidence for ammonium-bearing minerals on Ceres. *Science* **255**, 1551–1553 (1992).
26. Rivkin, A. S. *Observations of Main-belt Asteroids in the 3- μm Region*. Ph.D. dissertation, Univ. Arizona (1997).
27. Fanale, F. P. & Salvail, J. R. The water regime of asteroid 1 Ceres. *Icarus* **82**, 97–110 (1989).
28. Saint-Pe, O., Combes, M. & Rigaut, F. Ceres surface properties by high-resolution imaging from earth. *Icarus* **105**, 263–271 (1993).
29. Drummond, J. O., Fungate, R. Q. & Christou, J. G. Full adaptive optics images of asteroids Ceres and Vesta; rotational poles and triaxial ellipsoid dimensions. *Icarus* **132**, 80–99 (1998).
30. Pavlovsky, C. et al. *ACS Instrument Handbook, Version 5.0* (Space Telescope Science Institute, Baltimore, 2004).

Acknowledgements We thank B. Carcich and K. Consroe for technical assistance. J. Burns and J. Veverka provided discussions. This work was supported by NASA through the Space Telescope Science Institute, which is operated by the Association of Universities for Research and Astronomy, Inc. L.A.M., C.T.R. and M.V.S. were supported by NASA's Dawn Discovery mission. A review by S. Dermott improved the manuscript.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.C.T. (thomas@baritone.astro.cornell.edu).

Wave acceleration of electrons in the Van Allen radiation belts

Richard B. Horne¹, Richard M. Thorne², Yuri Y. Shprits², Nigel P. Meredith¹, Sarah A. Glauert¹, Andy J. Smith¹, Shrikanth G. Kanekal³, Daniel N. Baker³, Mark J. Engebretson⁴, Jennifer L. Posch⁴, Maria Spasojevic⁵, Umran S. Inan⁵, Jolene S. Pickett⁶ & Pierrette M. E. Decreau⁷

The Van Allen radiation belts¹ are two regions encircling the Earth in which energetic charged particles are trapped inside the Earth's magnetic field. Their properties vary according to solar activity^{2,3} and they represent a hazard to satellites and humans in space^{4,5}. An important challenge has been to explain how the charged particles within these belts are accelerated to very high energies of several million electron volts. Here we show, on the basis of the analysis of a rare event where the outer radiation belt was depleted and then re-formed closer to the Earth⁶, that the long established theory of acceleration by radial diffusion is inadequate; the electrons are accelerated more effectively by electromagnetic waves at frequencies of a few kilohertz. Wave acceleration can increase the electron flux by more than three orders of magnitude over the observed timescale of one to two days, more than sufficient to explain the new radiation belt. Wave acceleration could also be important for Jupiter, Saturn and other astrophysical objects with magnetic fields.

Inward radial diffusion has long been established as the leading acceleration mechanism for outer radiation belt electrons^{7,8}. According to the theory, electrons are diffused across the magnetic field by global scale fluctuations in the Earth's magnetic and electric fields at frequencies that closely match the electron drift frequencies of a few millihertz (mHz) around the Earth. The process occurs more rapidly when ultralow frequency (ULF) waves at a few millihertz are enhanced⁹. By conservation of the first adiabatic invariant (proportional to the square of the particle momentum transverse to the magnetic field divided by the magnetic field strength) the particles are accelerated if they are diffused towards the planet. However, it has recently been suggested^{10,11} that electrons can be accelerated efficiently by electromagnetic waves at frequencies of a few kilohertz (kHz) propagating in the whistler mode, and known as 'whistler mode chorus'¹². The rare and unusual events that occurred between 29 October and 4 November 2003, widely known as the Hallowe'en storms⁶, provide a unique set of conditions to test the two leading acceleration theories.

Figure 1a shows the electron particle flux at energies 2–6 MeV during the Hallowe'en storms. The data are presented as a function of L (where L is the distance from the centre of the Earth to the point where a magnetic field line crosses the equator, measured in Earth radii). Before the storms, the outer radiation belt is centred near $L = 3.5$. As the first storm begins on 29 October, there is a large depletion in the electron flux for $L > 3.0$, and a re-formation of the outer radiation belt near $L = 2.5$. The intensity of the new belt varies, but increases significantly from 1 November for a period of three or four days. The flux subsequently decays and the outer

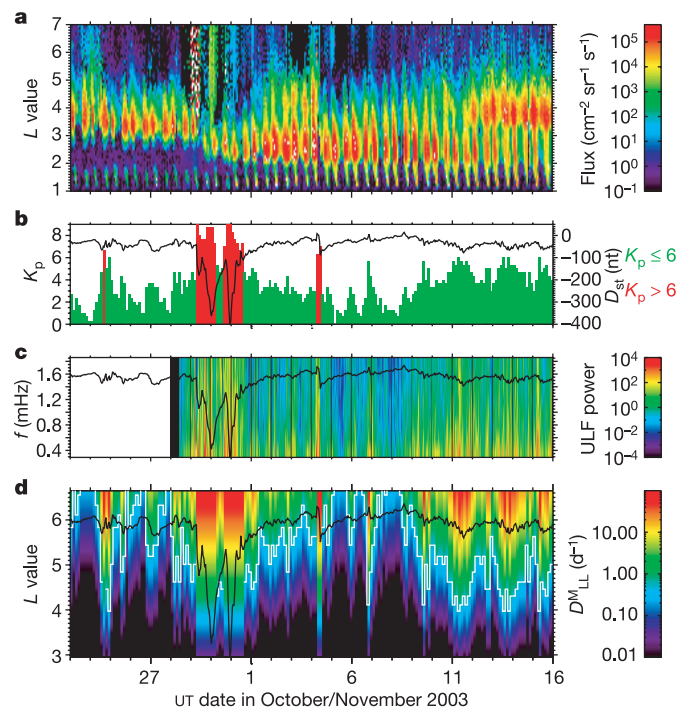


Figure 1 | Satellite and ground based data during the Hallowe'en storm.

a, Flux of 2–6 MeV electrons measured by the Solar Anomalous and Magnetospheric Particle Explorer (SAMPEX) satellite, colour coded according to the colour bar on the right. The data are averaged over each orbit. SAMPEX is a polar orbiting satellite; the flux modulations are due to the satellite passing over the South Atlantic Anomaly region where the outer radiation belt penetrates closer to the Earth. **b**, the K_p index (colour coded) and the D_{st} index (solid line). The former is a measure of the global disturbance in the Earth's magnetic field and is related to ULF waves, and the latter is a measure of electrical current systems and exhibits rapid negative excursions when geomagnetic storms are triggered. The Hallowe'en storms started on 29 October and ended by 4 November. The D_{st} index is also included in **c** and **d**. **c**, ULF wave power measured on the ground by Automatic Geophysical Observatory A80 in Antarctica at $L = 6.2$. The data are differenced to help identify signals and colour coded in $\text{nT}^2 \text{ Hz}$. **d**, Radial diffusion coefficient D_{LL}^M for magnetic field fluctuations, valid for $K_p < 6$. When $K_p > 6$, they are calculated for $K_p = 6$. The solid white line shows the L shell for which the radial diffusion coefficient is 0.5 d^{-1} .

¹British Antarctic Survey, Madingley Road, Cambridge CB3 0ET, UK. ²Department of Atmospheric and Oceanic Sciences, University of California Los Angeles, 405 Hilgard Avenue, Los Angeles, California 90095-1565, USA. ³Laboratory for Atmospheric and Space Physics, University of Colorado, 1234 Innovation Drive, Boulder, Colorado 80303-7814, USA. ⁴Department of Physics, Augsburg College, Minneapolis, Minnesota 55454, USA. ⁵STAR Laboratory, Stanford University, Stanford, California 94305, USA. ⁶Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242-1479, USA. ⁷LPCE, 3A Avenue de la recherche scientifique, 45071 Orleans, Cedex 2, France.

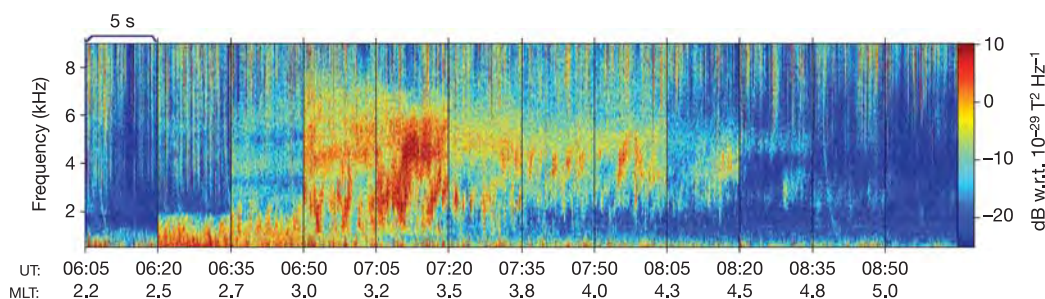


Figure 2 | Waves observed at Palmer station, Antarctica, on 1 November 2003. The power spectral density of electromagnetic waves (in the whistler mode) is shown in dB with respect to $10^{-29} \text{ T}^2 \text{ Hz}^{-1}$. Each panel is a five-second snapshot starting at the universal time (UT) indicated, with approximate magnetic local time (MLT) also shown. Wave power (identified as whistler mode chorus) is significantly enhanced between 2 kHz and 6 kHz on 1 November 2003 between 06:50 UT and 08:35 UT when the station was in

daylight, which causes strong attenuation of the waves as they pass through the upper atmosphere. Sunrise at Palmer at 100 km altitude is 05:00 UT on this day. The emissions faded owing to increased absorption as the Sun rises further, and possibly owing to drift of the source of low energy electrons out of the station viewing area. Strong waves were also detected at Halley Research station at $L = 4.3$ during the storms (not shown).

radiation belt is re-formed after 10 November at its usual location near $L = 4$.

A key feature of this event is that after 1 November the electron flux increases as the global level of magnetic activity, as measured by the K_p index (Fig. 1b), decreases and as conditions recover from the magnetic storms, as indicated by the D_{st} index returning to near zero (Fig. 1b). Data from one of the Automatic Geophysical Observatories, operated by the British Antarctic Survey in Antarctica, also show that the power level of ULF waves decreased as the flux increased (Fig. 1c). As ULF waves drive particle transport across the magnetic field, the data suggest that radial diffusion may not be the most effective process responsible for the flux increase.

The K_p index is often used as a proxy to determine the efficiency of inward radial diffusion¹³. Using this proxy, we have calculated the radial diffusion coefficient, D_{LL} , which determines the timescale for inward electron transport. The proxy is only valid for $K_p < 6$, and thus the results are not valid during 29, 30 and 31 October when the new belt began to form. Thus it is possible that inward radial diffusion could contribute to the formation of the new belt¹⁴ before 1 November. However, from 1 November onwards, the flux in the new belt increased by a factor of ~ 10 , but D_{LL} decreased, and was comparable to the values before the new belt (on 28 October) when there was no flux increase (compare white lines). In fact D_{LL} was even larger on 11 and 13 November when the new belt was decaying. Thus

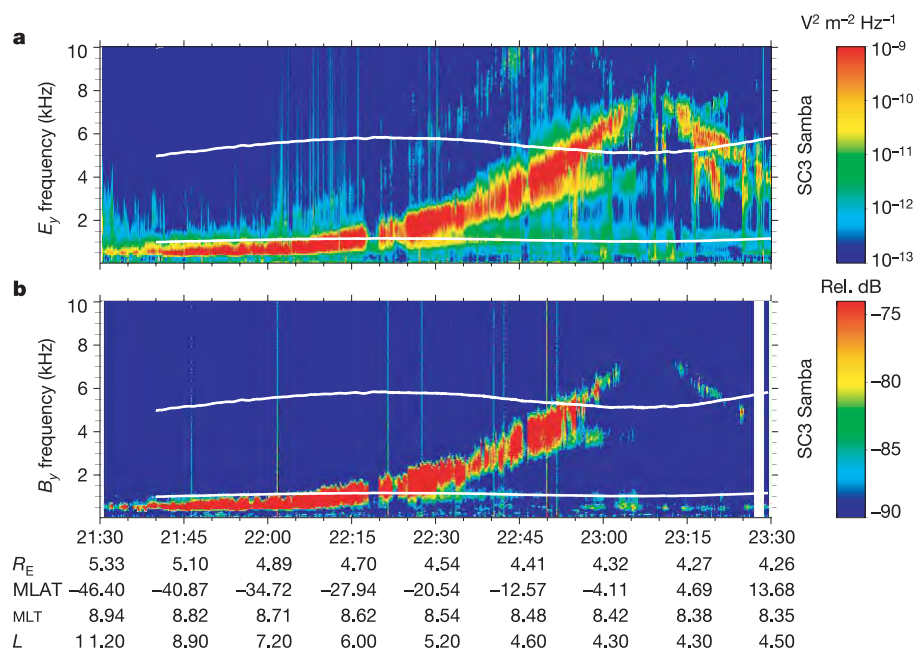


Figure 3 | Waves observed by the Cluster spacecraft. Data from the Wideband (WBD) plasma wave instrument on one of the four Cluster spacecraft (Samba) is shown for 31 October 2003. The wave electric field E_y , transverse to the ambient magnetic field is shown in **a**, and wave magnetic field B_y , in **b**. The colour bars on the right give the calibrated wave power spectral density for the wave electric field (**a**), and relative wave power spectral density (in dB) for the wave magnetic field (**b**). The instrument cycles between the electric and magnetic antennas recording data for 42 s (electric) and 10 s (magnetic). The data obtained from the electric (magnetic) antenna during each cycle is dilated to fill the data gap that is created when WBD is obtaining data from the magnetic (electric) antenna.

Both antennas detect chorus waves indicating that the waves are electromagnetic. Magnetic wave power (**b**) above 4 kHz is attenuated owing to low pass filtering in the search coils and is thus not representative of the true wave power. The white lines denote $0.5f_{ce}$ and $0.1f_{ce}$, where f_{ce} is the local electron cyclotron frequency ($f_{ce} = |e|B/(2\pi m_e)$ where $|e|$ is the electron charge, m_e is the electron mass and B is the ambient magnetic field obtained from the fluxgate magnetometer on board the spacecraft). At the magnetic equator f_{ce} is approximately 10 kHz, and the electron plasma frequency f_{pe} , obtained from the Whisper sounder, is approximately 35 kHz. At 22:46 UT the peak wave power spectral density is $\sim 6 \times 10^{-6} \text{ nT}^2 \text{ Hz}^{-1}$ at $\sim 0.35f_{ce}$, and tends to increase with magnetic latitude (MLAT).

inward radial diffusion alone cannot explain the increase in 2–6 MeV electron flux from 1 November onwards.

Electron acceleration by whistler mode chorus waves is the alternative leading theory^{10,11}. The waves are excited by an unstable anisotropic¹⁵ distribution of ~ 10 keV electrons, possibly via non-linear interactions¹⁶, and transfer the energy to accelerate a fraction of the electrons to very high energies. Acceleration occurs via a cyclotron resonance where the wave frequency is Doppler shifted to the cyclotron frequency of the particles. For a broad band of waves, typically observed inside the Earth's magnetic field, resonance can extend from energies of ~ 10 keV up to several MeV (ref. 11). Such local acceleration is most effective in regions where the ratio of the electron plasma to cyclotron frequency (f_{pe}/f_{ce}) is low¹⁷, typically < 4 . Under normal conditions, the plasma density is high and f_{pe}/f_{ce} is large near $L = 2.5$, and hence wave acceleration is not expected. However, during the Hallowe'en storms the region of high density was confined to $L < 2.0$ on 31 October, and remained inside $L = 2.5$ between 06–12 magnetic local time (MLT) until about 4 November⁶. Using a dipole magnetic field, and a density model appropriate to this region¹⁸, we find $f_{pe}/f_{ce} \approx 2.5$, which suggests wave acceleration could be important.

Unfortunately, there are no direct wave observations in the new radiation belt near $L = 2.5$ during this event. However, whistler mode chorus waves can be guided along the magnetic field, and were detected at Palmer station, Antarctica, at $L = 2.4$ on 1 November (Fig. 2). Chorus waves are only observed at Palmer during very large

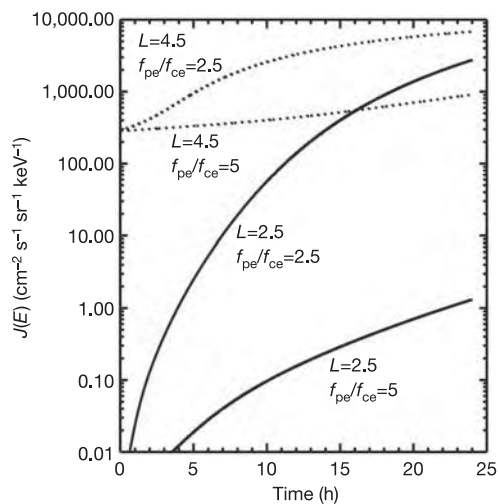


Figure 4 | Simulation results. Electron flux at 1 MeV due to wave acceleration for $L = 2.5$ (solid lines) and $L = 4.5$ (dotted lines). The flux was obtained by solving a one-dimensional Fokker-Planck equation given by equation (9) in ref. 20, and includes losses due to pitch angle diffusion. It is assumed that the electron pitch angle distribution is isotropic. The diffusion rates (bounced-averaged pitch angle and energy) corresponding to a peak power spectral density of $\sim 2 \times 10^{-6} \text{ nT}^2 \text{ Hz}^{-1}$ are given in Supplementary Table 1. The diffusion rates were scaled down by a factor of 4 (to represent the fraction of MLT over which the waves are present), and interpolated onto a high-resolution grid. The loss timescale was assumed to be equal to the inverse of the pitch angle diffusion rate at the edge of the loss cone. Acceleration rates exceed loss rates for $E > 300$ keV (ref. 20). Since SAMPEX did not measure electrons below 2 MeV, the initial flux was obtained by averaging CRRES satellite data for low magnetic activity (using auroral electrojet index, $AE < 100$ nT) at $2.4 < L < 2.6$ and $4.4 < L < 4.6$ for energies between 153 keV and 1.58 MeV. At $L = 2.5$ the flux for $E > 1.09$ MeV was at the noise level ($10^{-2} \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ keV}^{-1}$) and set to zero in the model. The flux was interpolated onto a high-resolution grid, and converted into phase space density. Fixed boundary conditions were used with the flux set to the initial value at 153 keV, and zero at 11 MeV. The code calculates phase space density, and converts the results back into flux for comparison with data.

storms and so the observation, particularly during sunlight, is exceptional. Similar waves were observed on 29, 30 and 31 October 2003. Since chorus waves are generated in space, in low-density regions¹⁹, the ground observations provide strong evidence for waves in space near the equator.

Strong whistler mode chorus waves were also detected at larger L ($L > 4.3$) on 31 October by the Cluster satellites (Fig. 3). The active sounder experiment on the satellite also provided an accurate measurement of $f_{pe}/f_{ce} = 3.5$ at the equator, and found that f_{pe}/f_{ce} decreased with increasing latitude, providing the right conditions for wave acceleration.

Using a one-dimensional Fokker-Planck equation²⁰, we have calculated the increase in electron flux due to whistler mode chorus waves (Fig. 4). Loss and acceleration due to chorus waves are included via diffusion rates calculated from the PADIE code²¹ (see Supplementary Data). Since chorus waves are usually present outside the high-density region during magnetic storms¹⁹, we assume that the wave power spectral density observed by Cluster at $L = 4.3$ are also typical of $L = 2.5$. We also assume that the waves only exist for 6 h out of 24 h of MLT, as shown by the statistical distribution of wave occurrence¹⁹.

At $L = 2.5$, for $f_{pe}/f_{ce} = 2.5$, the electron flux increases from the noise level and exceeds the average flux in the heart of the outer radiation belt ($L = 4.5$) within 24 h. This is more than sufficient to account for the enhancement in the new radiation belt after 1 November. For comparison, at $L = 4.5$ the flux increases by a factor of ~ 20 for $f_{pe}/f_{ce} = 2.5$. To illustrate the sensitivity of the results, the flux increase is also shown for $f_{pe}/f_{ce} = 5$. Although the flux increases at $L = 2.5$ by a factor of ~ 100 within 24 h, it remains below detectable levels. Thus wave acceleration is most effective owing to the low density at $L = 2.5$ as a result of the confinement of the high-density region to $L < 2.5$ during this event⁹. Losses are not significant, since low values of f_{pe}/f_{ce} increase the phase velocity of the waves so that energy diffusion is more effective than pitch angle diffusion, particularly for energies > 300 keV.

After 1 November the flux in the new radiation belt increased from a level that had already been enhanced. To represent a higher initial level, we repeated the calculation for a minimum flux level of $1 \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ keV}^{-1}$, and then $10^2 \text{ cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ keV}^{-1}$. After 24 h, the flux at $L = 2.5$, $f_{pe}/f_{ce} = 2.5$, reached approximately the same level as shown in Fig. 4.

The results suggest that wave-particle interactions not only play an important role determining the structure of the radiation belts via losses to the atmosphere²², but also are a principal mechanism for their formation in low-density regions. The theory is applicable to particle acceleration at Jupiter and Saturn, and to other astrophysical objects with magnetic fields.

Received 27 April; accepted 10 June 2005.

1. Van Allen, J. A. in *Discovery of the Magnetosphere* (eds Gillmor, C. S. & Spreiter, J. R.) 235–251 (Vol. 7, History of Geophysics, American Geophysical Union, Washington DC, 1997).
2. Baker, D. N., Blake, J. B., Klebesadel, R. W. & Higbie, P. R. Highly relativistic electrons in the Earth's outer magnetosphere 1. Lifetimes and temporal history 1979–1984. *J. Geophys. Res.* **91**, 4265–4276 (1986).
3. Li, X., Baker, D. N., Kanekal, S. G., Looper, M. & Temerin, M. Long term measurements of radiation belts by SAMPEX and their variations. *Geophys. Res. Lett.* **28**, 3827–3830 (2001).
4. Baker, D. N., Allen, J. H., Kanekal, S. G. & Reeves, G. D. Disturbed space environment may have been related to pager satellite failure. *Eos* **79**, 477 (1998).
5. Webb, D. F. & Allen, J. H. Spacecraft and ground anomalies related to the October–November 2003 solar activity. *Space Weath.* **2**, doi:10.1029/2004SW000075 (2004).
6. Baker, D. N. *et al.* An extreme distortion of the Van Allen belt arising from the Hallowe'en solar storm in 2003. *Nature* **432**, 878–881 (2004).
7. Falthammar, C.-G. Effects of time dependent electric fields on geomagnetically trapped radiation. *J. Geophys. Res.* **70**, 2503–2516 (1965).
8. Schulz, M. & Lanzerotti, L. J. *Particle Diffusion in the Radiation Belts* (Springer, New York, 1974).

9. Elkington, S. R., Hudson, M. K. & Chan, A. A. Acceleration of relativistic electrons via drift resonant interactions with toroidal-mode Pc-5 ULF oscillations. *Geophys. Res. Lett.* **26**, 3273–3276 (1999).
10. Summers, D., Thorne, R. M. & Xiao, F. Relativistic theory of wave-particle resonant diffusion with application to electron acceleration in the magnetosphere. *J. Geophys. Res.* **103**, 20487–20500 (1998).
11. Horne, R. B. & Thorne, R. M. Potential waves for relativistic electron scattering and stochastic acceleration during magnetic storms. *Geophys. Res. Lett.* **25**, 3011–3014 (1998).
12. Tsurutani, B. T. & Smith, E. J. Postmidnight chorus: A substorm phenomenon. *J. Geophys. Res.* **79**, 118–127 (1974).
13. Brautigam, D. H. & Albert, J. M. Radial diffusion analysis of outer radiation belt electrons during the October 9, 1990, magnetic storm. *J. Geophys. Res.* **105**, 291–309 (2000).
14. Shprits, Y. Y. & Thorne, R. M. Time dependent radial diffusion modeling of relativistic electrons with realistic loss rates. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL019591 (2004).
15. Kennel, C. F. & Petschek, H. E. Limit on stably trapped particle fluxes. *J. Geophys. Res.* **71**, 1–28 (1966).
16. Helliwell, R. A. A theory of discrete emissions from the magnetosphere. *J. Geophys. Res.* **72**, 4773–4790 (1967).
17. Horne, R. B., Glauert, S. A. & Thorne, R. M. Resonant diffusion of radiation belt electrons by whistler-mode chorus. *Geophys. Res. Lett.* **30**, doi:10.1029/2003GL016963 (2003).
18. Sheeley, B. W., Moldwin, M. B., Rassoul, H. K. & Anderson, R. R. An empirical plasmasphere and trough density model: CRRES observations. *J. Geophys. Res.* **106**, 25631–25641 (2001).
19. Meredith, N. P., Horne, R. B., Thorne, R. M. & Anderson, R. R. Favoured regions for chorus-driven electron acceleration to relativistic energies in the Earth's outer radiation belt. *Geophys. Res. Lett.* **30**, 1871, doi:10.1029/2003GL017698 (2003).
20. Horne, R. B. *et al.* Timescale for radiation belt electron acceleration by whistler mode chorus waves. *J. Geophys. Res.* **110**, A03225, doi:10.1029/2004JA010811 (2005).
21. Glauert, S. A. & Horne, R. B. Calculation of pitch angle and energy diffusion coefficients with the PADIE code. *J. Geophys. Res.* **110**, A04206, doi:10.1029/2004JA010851 (2005).
22. Lyons, L. R. & Thorne, R. M. Equilibrium structure of radiation belt electrons. *J. Geophys. Res.* **78**, 2142–2149 (1973).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank E. Lucek for providing fluxgate magnetometer data from the Cluster spacecraft, and N. Cornilleau-Wehrin for an independent assessment of the wave magnetic power spectral density. This work was supported in part by the UK Natural Environment Research Council (NERC), the NSF and NASA.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.B.H. (R.Horne@bas.ac.uk).

Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies

D. Collin^{1*}, F. Ritort^{2*}, C. Jarzynski³, S. B. Smith⁴, I. Tinoco Jr⁵ & C. Bustamante^{4,6}

Atomic force microscopes and optical tweezers are widely used to probe the mechanical properties of individual molecules and molecular interactions, by exerting mechanical forces that induce transitions such as unfolding or dissociation. These transitions often occur under nonequilibrium conditions and are associated with hysteresis effects—features usually taken to preclude the extraction of equilibrium information from the experimental data. But fluctuation theorems^{1–5} allow us to relate the work along nonequilibrium trajectories to thermodynamic free-energy differences. They have been shown to be applicable to single-molecule force measurements⁶ and have already provided information on the folding free energy of a RNA hairpin^{7,8}. Here we show that the Crooks fluctuation theorem⁹ can be used to determine folding free energies for folding and unfolding processes occurring in weak as well as strong nonequilibrium regimes, thereby providing a test of its validity under such conditions. We use optical tweezers¹⁰ to measure repeatedly the mechanical work associated with the unfolding and refolding of a small RNA hairpin¹¹ and an RNA three-helix junction¹². The resultant work distributions are then analysed according to the theorem and allow us to determine the difference in folding free energy between an RNA molecule and a mutant differing only by one base pair, and the thermodynamic stabilizing effect of magnesium ions on the RNA structure.

The Crooks fluctuation theorem⁹ (CFT) predicts a symmetry relation in the work fluctuations associated with the forward and reverse changes a system undergoes as it is driven away from thermal equilibrium by the action of an external perturbation. This theorem applies to processes that are microscopically reversible, and its experimental evaluation in small systems is crucial to understand better the foundations of nonequilibrium physics¹³. A consequence of the CFT is Jarzynski's equality¹⁴, which relates the equilibrium free-energy difference ΔG between two equilibrium states to an exponential average (denoted by angle brackets) of the work done on the system, W , taken over an infinite number of repeated nonequilibrium experiments, $\exp(-\Delta G/k_B T) = \langle \exp(-W/k_B T) \rangle$. The equality has been developed⁶ into a formalism that allows us to use nonequilibrium single-molecule pulling experiments to reconstruct free-energy profiles or potentials of mean force¹⁵ along reaction coordinates. Experimental testing of Jarzynski's equality in single-molecule force experiments¹⁶ used the P5ab RNA hairpin^{7,8}, which can be folded and unfolded quasi-reversibly. But for processes that occur far from equilibrium, the applicability of Jarzynski's equality is hampered by large statistical uncertainties that arise from the sensitivity of the exponential average to rare events^{17,18} (low values of W). Moreover, although the equality $\langle W \rangle = \Delta G$ holds for processes occurring near equilibrium, spatial drift in the experimental

system usually makes it difficult in practice to extract unfolding free energies using small loading rates (below a few pN s⁻¹). Drift effects decrease noticeably for larger pulling speeds, making it possible to obtain more reliable experimental data (and also good statistics as a large number of pulls can be executed in a reasonable time), but at the expense of a more irreversible unfolding process. Here we show that significant improvements can be obtained by using the CFT, which provides a more robust and more rapidly converging method to extract equilibrium free energies from non-equilibrium processes.

The CFT allows us to quantify the amount of hysteresis observed in the values of the irreversible work done to unfold and refold a macromolecule. Let $P_U(W)$ denote the probability distribution of the values of the work performed on the molecule in an infinite number of pulling experiments along the unfolding (U) process, and define $P_R(W)$ analogously for the reverse (R) process. For the CFT to be applicable, the unfolding and refolding processes need to be related by time-reversal symmetry, that is, in our experiments, the optical trap used to manipulate the molecule must be moved at the same speeds during unfolding and refolding. Moreover, the molecular transition probed always has to start in an equilibrium state (folded in the unfolding process, and denatured or unfolded in the refolding process) and reach a well-defined final state. The CFT⁹ then predicts that:

$$\frac{P_U(W)}{P_R(-W)} = \exp\left(\frac{W - \Delta G}{k_B T}\right) \quad (1)$$

where ΔG is the free-energy change between the final and the initial states, and thus equal to the reversible work associated with this process. Note that the CFT does not require that the system studied reaches its final equilibrium state immediately after the unfolding and refolding processes have been completed; it is only the control parameter that needs to attain its final value, whereas the system may continue to equilibrate to a well-defined state that is consistent with the final value of the control parameter. The equilibration occurs without change of the control parameter, and therefore contributes no work. In principle, W is an integral over the external variation of a control parameter⁹, for example, the position of the optical trap or the time⁶. In our experiments, W is well approximated by the familiar force-versus-extension integral:

$$W = \sum_{i=1}^{N_s} F_i \Delta x_i \quad (2)$$

where x_i is the distance between the ends of the molecule and N_s is the number of intervals used in the sum (see ref. 6 for a thorough discussion of this issue). Relation (1) quantifies hysteretic effects in the pulling experiment: work values larger than ΔG occur most often

¹Merck & Co. Inc., Automated Biotechnology Department, North Wales, Pennsylvania 19454, USA. ²Departament de Física Fonamental, Facultat de Física, Universitat de Barcelona, Diagonal 647, 08028 Barcelona, Spain. ³T-13 Complex Systems, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. ⁴Howard Hughes Medical Institute, ⁵Department of Chemistry, ⁶Departments of Physics and Molecular & Cell Biology, University of California, Berkeley, California 94720, USA.

*These authors contributed equally to this work.

along the unfolding path while (absolute) values smaller than ΔG occur more often along the refolding path. As can be seen from equation (1), the CFT states that although $P_U(W)$, $P_R(-W)$ depend on the pulling protocol, their ratio depends only on the value of ΔG . Thus the value of ΔG can be determined once the two distributions are known. In particular, the two distributions cross at $W = \Delta G$:

$$P_U(W) = P_R(-W) \Rightarrow W = \Delta G \quad (3)$$

regardless of the pulling speed. Although the simple identity (3) already gives an estimate of ΔG , it is not necessarily very precise because it uses only the local behaviour of the distribution around $W = \Delta G$. Using the whole work distribution increases the precision of the free-energy estimate¹⁹. In particular, as we show below, when the overlapping region of work values between the unfolding and refolding work distributions is too narrow (as may happen for large values of the average dissipated work, defined as $\langle W_{\text{dis}} \rangle = \langle W \rangle - \Delta G$), the use of Bennett's acceptance ratio method²⁰ makes it possible to extract accurate estimates of ΔG using the CFT (see the Supplementary Information).

We first experimentally test the validity of the CFT for a molecular transition occurring near equilibrium. For this, we use a short interfering (si)RNA hairpin that targets the messenger RNA of the CD4 receptor of the human immunodeficiency virus (HIV)¹¹ and that unfolds irreversibly but not too far from equilibrium at accessible experimental pulling speeds (dissipated work values less than $6k_B T$). Under these conditions, the unfolding and refolding work distributions overlap over a sufficiently large range of work values to justify the use of the direct method to experimentally test equation (1). The work done on the molecules during either pulling or relaxation is given by the areas below the corresponding force–extension curves (Fig. 1).

Unfolding and refolding work distributions at three different pulling speeds are shown in Fig. 2. Irreversibility increases with the pulling speed and unfolding–refolding work distributions become progressively more separated. Note, however, that the unfolding and the refolding distributions cross at a value of the work $\Delta G = 110.3 \pm 0.5 k_B T$ that does not depend on the pulling speed, as predicted by equation (3). Moreover, the work distributions also

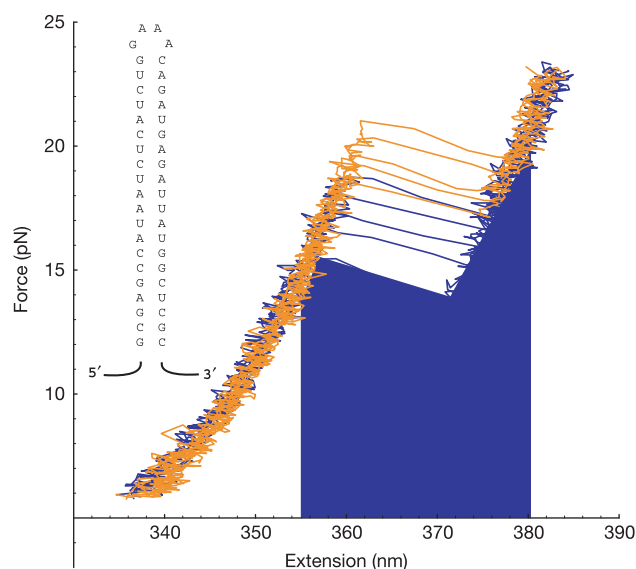


Figure 1 | Force–extension curves. The stochasticity of the unfolding and refolding process is characterized by a distribution of unfolding or refolding work trajectories. Five unfolding (orange) and refolding (blue) force–extension curves for the RNA hairpin are shown (loading rate of 7.5 pN s^{-1}). The blue area under the curve represents the work returned to the machine as the molecule switches from the unfolded to the folded state. The RNA sequence is shown as an inset.

satisfy the CFT, that is, equation (1) (see the Supplementary Information). We also notice that work distributions are compatible with, and can be fitted to, gaussian distributions (data not shown). After subtracting the contribution arising from the entropy loss due to the stretching of the molecular handles attached on both sides of the hairpin ($\Delta G^{\text{handles}} = 23.8 k_B T$) and of the extended single-stranded (ss)RNA ($\Delta G^{\text{ssRNA}} = 23.7 \pm 1 k_B T$) from the total work, $\Delta G^{\text{exp}} = 110.3 \pm 0.5 k_B T$, we obtain for the free energy of unfolding at zero force $\Delta G_0^{\text{exp}} = 62.8 \pm 1.5 k_B T = 37.2 \pm 1 \text{ kcal mol}^{-1}$ (at 25°C , in 100 mM Tris-HCl, pH 8.1, 1 mM EDTA), in excellent agreement with the result obtained using the Visual OMP from DNA software²¹ $\Delta G_0^{\text{fold}} = 38 \text{ kcal mol}^{-1}$ (at 25°C , in 100 mM NaCl).

To extend the experimental test of the validity of the CFT to the very-far-from-equilibrium regime where the work distributions are no longer gaussian, we apply the CFT to determine: (1) the difference in folding free energy between an RNA molecule and a mutant that differs only by one base-pair, and (2) the thermodynamic stabilizing effect of Mg^{2+} ions on the RNA structure. The RNA we consider is a three-helix junction of the 16S ribosomal RNA of *Escherichia coli*¹² that binds the S15 protein. The secondary structure of this RNA is a common feature in RNA structures^{22–24} that plays, in this case, a crucial role in the folding of the central domain of the 30S ribosomal subunit. For comparison, and to verify the accuracy of the method, we have pulled the wild type and a C-G to G-C mutant (C754G to G587C) of the three-helix junction.

Figure 3 depicts the unfolding and refolding work distributions for the wild-type and mutant molecules (work values were binned into about 10–20 equally spaced intervals). For both molecules, the distributions display a very narrow overlapping region. In contrast with the hairpin distribution, the average dissipated work for the unfolding pathway is now much larger—in the range $20\text{--}40 k_B T$ —and the unfolding work distribution shows a large tail and strong deviations from gaussian behaviour. Thus, these molecules are ideal to test the validity of equation (1) in the far-from-equilibrium regime. As shown in the inset of Fig. 3, the plot of the log ratio of the unfolding to the refolding probabilities versus total work done on the molecule can be fitted to a straight line with a slope of 1.06, thus establishing the validity of the CFT (see equation (1)) under far-from-equilibrium conditions. Our measurements reveal the presence of long tails in the work distribution $P_U(W)$ along the unfolding path

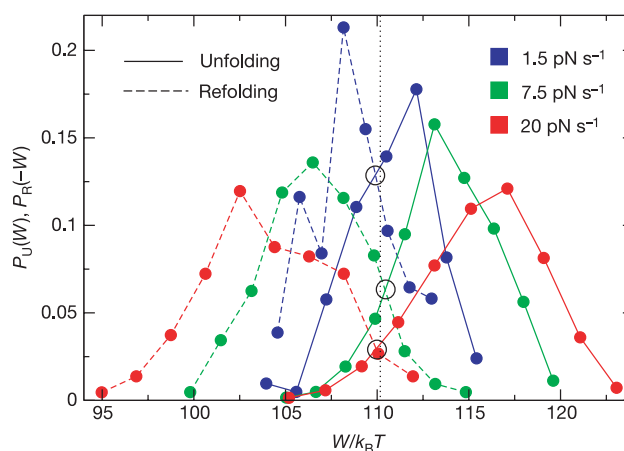


Figure 2 | Test of the CFT using an RNA hairpin. Work distributions for RNA unfolding (continuous lines) and refolding (dashed lines). We plot negative work, $P_R(-W)$, for refolding. Statistics: 130 pulls and three molecules ($r = 1.5 \text{ pN s}^{-1}$), 380 pulls and four molecules ($r = 7.5 \text{ pN s}^{-1}$), 700 pulls and three molecules ($r = 20.0 \text{ pN s}^{-1}$), for a total of ten separate experiments. Good reproducibility was obtained among molecules (see Supplementary Fig. S2). Work values were binned into about ten equally spaced intervals. Unfolding and refolding distributions at different speeds show a common crossing around $\Delta G = 110.3 k_B T$.

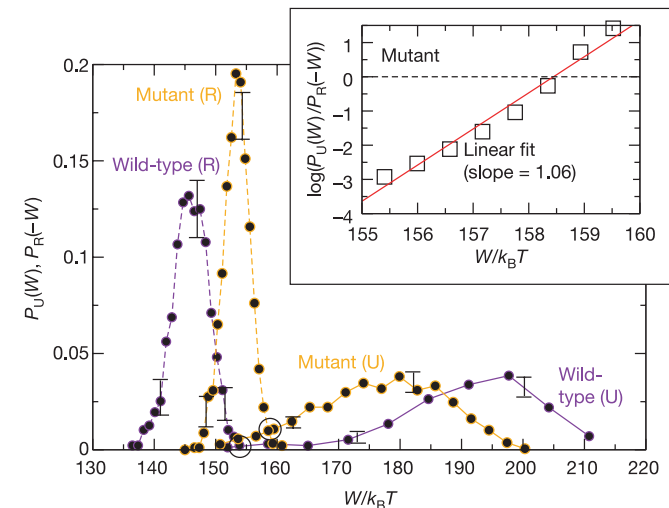


Figure 3 | Free-energy recovery and test of the CFT for non-gaussian work distributions. Experiments were carried out on the wild-type and mutant S15 three-helix junction without Mg^{2+} . Unfolding (continuous lines) and refolding (dashed lines) work distributions. Statistics: 900 pulls and two molecules (wild type, purple); 1,200 pulls and five molecules (mutant type, orange). Crossings between distributions are indicated by black circles. Work histograms were found to be reproducible among different molecules (error bars indicating the range of variability). Inset, test of the CFT for the mutant. Data have been linearly interpolated between contiguous bins of the unfolding and refolding work distributions.

and narrow work distributions $P_R(W)$ along the refolding path. These distributions complement each other, one being large where the other is small, thereby providing thermodynamically important information about the free-energy landscape.

Bennett's acceptance ratio method gives $\Delta G^{\text{exp}} = 154.1 \pm 0.4 k_B T$ and $\Delta G^{\text{exp}} = 157.9 \pm 0.2 k_B T$ for unfolding the wild-type and mutant types, respectively, giving a difference between the two forms $\Delta \Delta G_0^{\text{exp}} = \Delta \Delta G^{\text{exp}} = 3.8 \pm 0.6 k_B T$. After subtracting the (identical for both molecules) handle and RNA entropy loss contributions ($97 \pm 1 k_B T$) we get $\Delta G_0^{\text{exp}} = 57 \pm 1.5 k_B T$ (wild type) and $\Delta G_0^{\text{exp}} = 60.8 \pm 1.5 k_B T$ (mutant), the error increasing owing to the uncertainty in the contributions coming from the stretching of ssRNA. Free-energy prediction programs such as Mfold²⁵ and Visual OMP²¹ give a $\Delta \Delta G_0^{\text{fold}} = 2 \pm 2 k_B T$ between the forms (at 25 °C and 100 mM NaCl). Thus, when combined with acceptance ratio methods, the CFT furnishes a method precise enough to determine the difference in the folding free energies of RNA molecules differing only by one base pair in 34 base pairs.

Finally we apply equation (1) to obtain the free energy of

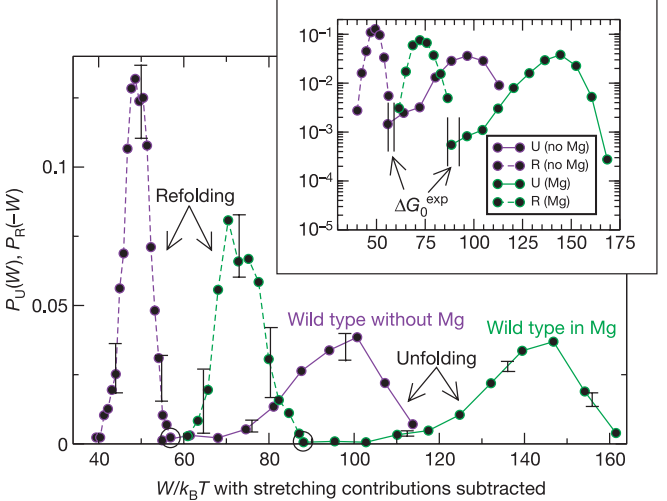


Figure 4 | Use of CFT to extract the stabilizing contribution of Mg^{2+} to the free energy of the S15 three-helix junction (wild type). Unfolding (continuous lines) and refolding (dashed lines) work distributions. Green curves, 450 pulls and two molecules in Mg^{2+} ; purple curves, 900 pulls and two molecules without Mg^{2+} . Crossings between distributions are indicated by black circles. Work histograms are reproducible between the molecules (error bars indicating the range of variability). Inset, the same histograms in logarithmic scale (axes labels as for the main panel) showing (vertical black bars) the regions of work values where unfolding and refolding distributions are expected to cross each other by Bennett's acceptance ratio method (Supplementary Information).

stabilization by Mg^{2+} of the S15 three-helix junction. These values are often difficult to access using bulk methods because melting temperatures of tertiary folded RNAs are frequently higher than the boiling point of water, and Mg^{2+} catalyses the hydrolysis of RNA at increased temperatures²⁶. Figure 4 depicts the work histograms in the presence and absence of Mg^{2+} (at constant ionic strength); stretching contributions differ in the presence and absence of magnesium ions ($116.8 k_B T$ and $97 k_B T$, respectively). These values have been subtracted from the work data to properly compare the unfolding free energies of both molecules. The strong increase of irreversibility due to Mg^{2+} can be seen in the large value of the average dissipated work (about $50 k_B T$ along the unfolding reaction and $16 k_B T$ along the refolding path). Applying Bennett's acceptance ratio method for the molecule in the presence of magnesium yields $\Delta G^{\text{exp}} = 205.5 \pm 1.5 k_B T$ and (after subtracting the stretching contributions) gives $\Delta G_0^{\text{exp}} = 88.7 \pm 2.5 k_B T$ for the unfolding reaction of the wild-type junction in 4 mM $MgCl_2$. The difference in free energies of unfolding in the presence and absence of Mg^{2+} , $\Delta \Delta G_0^{\text{exp}} = -31.7 \pm 2 k_B T$, gives the free energy of stabilization associated with the binding of Mg^{2+}

Table 1 | Summary of results obtained for all molecules

Molecule	W_m^U	W_m^R	σ_U	σ_R	$W_{\text{cum}}^{(2)}$	W_J^U	W_J^R	W_J^{est}	ΔG^{exp}	ΔG_0^{exp}	W_{dis}^U	W_{dis}^R	R_U	R_R
Hairpin (1.5 pN s ⁻¹)*	110.9	108.7	2.35	2.21	109.7 (0.2)	107.4 (0.7)	110.9 (0.2)	109.1 (0.5)	110.0 (0.2)	62.5 (1.2)	0.9	1.3	3.1	1.9
Hairpin (7.5 pN s ⁻¹)	113.8	106.6	2.63	2.84	110.3 (0.2)	109.7 (0.7)	110.9 (0.5)	110.3 (0.5)	110.3 (0.5)	62.8 (1.5)	3.5	3.7	0.98	1.10
Hairpin (20 pN s ⁻¹)	115.7	104.1	3.2	3.5	110.1 (0.2)	110.2 (0.7)	108.6 (0.2)	109.4 (0.4)	110.2 (0.6)	62.9 (1.6)	5.4	6.2	0.94	0.98
S15 (wild, no Mg)	191.3	145.9	11.3	2.9	158.7 (0.8)	155.2 (1.4)	149.3 (0.2)	152.2 (0.7)	154.1 (0.4)	57.0 (1.5)	36.3	9.1	1.75	0.46
S15 (mutant, no Mg)	176.5	153.4	10.6	2.1	156.0 (0.4)	152.4 (5.0)	155.7 (0.2)	154.1 (0.3)	157.9 (0.2)	60.8 (1.5)	18.6	4.5	3.02	0.49
S15 (wild, Mg)	256.4	190.3	12.2	5.0	213.0 (1.3)	207.0 (4.0)	199.8 (0.6)	203.6 (2.0)	205.5 (1.5)	88.7 (2.5)	50.9	15.2	1.46	0.82

W_m^U and W_m^R , σ_U and σ_R , W_J^U and W_J^R are the average total work, standard deviations and predictions obtained by using Jarzynski's equality along the unfolding (U) and refolding (R) paths. $W_{\text{cum}}^{(2)}$ is the estimate obtained by Hummer³⁰, $W_{\text{cum}}^{(2)} = (W_m^U + W_m^R)/2 - (\sigma_U^2 - \sigma_R^2)/12 k_B T$, which gives the leading correction to the linear response prediction, W_J^{est} is the average of the estimates obtained by using Jarzynski's equality along the unfolding and refolding paths $W_J^{\text{est}} = (W_J^U + W_J^R)/2$. ΔG^{exp} is our best estimate obtained by using the acceptance ratio method; ΔG_0^{exp} is the final estimate for the unfolding free energy at zero force after subtracting the handles contribution, $W_{\text{dis}}^U = |W_J^U - \Delta G^{\text{exp}}|$ is the average dissipated work (for the analysis of the hairpin data we took $\Delta G^{\text{exp}} = 110.3$ for all pulling rates) and $R_{U,R} = \frac{\sigma_{U,R}^2}{2k_B T W_{\text{dis}}^U}$ is a parameter that is equal to 1 for gaussian work distributions⁸. Statistical errors are given for Jarzynski's equality and the crossing estimates. These were obtained using the bootstrap method. All work values (except ΔG_0^{exp}) include the handle and RNA stretching contributions and are given in units of $k_B T$ at $T = 298 K$. In parentheses we indicate the errors in units of $k_B T$.

*Data for the RNA hairpin at 1.5 pN s⁻¹ are also included for completion; however, at such low loading rates drift effects are very large and data are very noisy as revealed by the values of R_U and R_R , which differ too much from 1.

ions to the S15 three-helix junction through both specific and non-specific (shielding) interactions.

These results illustrate that when used in conjunction with an appropriate fluctuation theorem, nonequilibrium single-molecule force measurements can provide equilibrium information such as folding free energies, even if the process studied occurs under far-from-equilibrium conditions. The approach works using soft optical traps but is probably limited to processes that dissipate less than $100 k_B T$. Whether it can be extended to studies that use much stiffer atomic-force-microscope cantilevers to pull proteins²⁷ is at present being examined in our laboratory. Finally, the initial and final states of molecular interactions such as ligand binding or macromolecular assembly usually do not correspond to measurable molecular extensions; in such cases, the approach as described cannot at present be applied.

METHODS

Sample preparation. The RNA molecules were prepared as previously described by ref. 16. Each DNA sequence corresponding to the three different RNAs was cloned separately into pBR322 vector between *EcoRI* and *HindIII* sites. A polymerase chain reaction (PCR) was used to amplify a DNA sequence containing an upstream T7 promoter, the RNA sequence of interest and flanking DNA sequences corresponding to the 'handles'. The handles correspond to a sequence of pBR322 (NCBI ID 'J01749') from nucleotides 3838 to 1 and from 29 to 629, respectively. The three RNA sequences were transcribed *in vitro* using T7 RNA polymerase²⁸. Two DNA handles were synthesized by PCR. The DNA handle upstream of the RNA was biotinylated at the 3'-end, whereas a digoxigenin moiety was attached to the 5'-end of the other handle. The RNA and two DNA handles were annealed by heating samples to 85 °C, followed by a slow cooling down to room temperature. The RNA hairpin was pulled in 100 mM Tris-HCl, pH 8.1, 1 mM EDTA buffer. The S15 three-helix junction and the mutant have been pulled in 62 mM KCl, 10 mM HEPES pH 7.8 buffer. In 4 mM MgCl₂ the KCl concentration was adjusted to 50 mM to work at the same ionic strength as in the absence of Mg²⁺.

Work measurements. Work probability distributions were obtained from many force-extension curves for a given molecule and aligned to a worm-like chain curve that best fitted the force-extension data at forces below the range of forces where the molecule unfolds. This procedure minimizes the effect of machine drift on the measured work values. For the worm-like chain fits we used $P \approx 10$ nm and $P \approx 1$ nm for the persistence lengths of the DNA/RNA hybrid handles and ssRNA respectively. Work values were integrated along the range of extension: [355 nm, 380 nm] for the hairpin; [326 nm, 392 nm] for the S15 three-helix junction without magnesium (wild and mutant); and [337 nm, 398 nm] for the S15 three-helix junction with magnesium. The free-energy contributions from stretching the handles and the ssRNA were then obtained by numerical integration of the worm-like chain reference curves using the values for the persistence and contour lengths of the polymers. To estimate the free energy of unfolding at zero, ΔG_0^{exp} , we subtract the free-energy contribution of the hybrid handles and ssRNA from the total reversible work across the transition, ΔG^{exp} , by using the expression^{7,29}: $\Delta G_0^{\text{exp}} = \Delta G^{\text{exp}} - \Delta G^{\text{handles}} - \Delta G^{\text{ssRNA}}$, where $\Delta G^{\text{handles}}$, ΔG^{ssRNA} are the entropy loss contributions due to the stretching of the molecular handles attached on both sides of the hairpin and of the extended ssRNA, respectively.

Received 11 April; accepted 2 July 2005.

- Evans, D. J., Cohen, E. G. & Morriss, G. P. Probability of second law violations in shearing steady states. *Phys. Rev. Lett.* **71**, 2401–2404 (1993).
- Gallavotti, G. & Cohen, E. G. D. Dynamical ensembles in nonequilibrium statistical mechanics. *Phys. Rev. Lett.* **74**, 2694–2697 (1995).
- Ciliberto, S. & Laroche, C. An experimental test of the Gallavotti-Cohen fluctuation theorem. *J. Phys. IV* **8** (Proc. 6), 215–220 (1998).
- Evans, D. J. & Searles, D. J. The fluctuation theorem. *Adv. Phys.* **51**, 1529–1585 (2002).
- Wang, G. M., Sevick, E. M., Mittag, E., Searles, D. J. & Evans, D. J. Experimental demonstration of violations of the second law of thermodynamics for small systems and short timescales. *Phys. Rev. Lett.* **89**, 050601 (2002).
- Hummer, G. & Szabo, A. Free-energy reconstruction from nonequilibrium

- single molecule experiments. *Proc. Natl Acad. Sci. USA* **98**, 3658–3661 (2001).
- Liphardt, J., Dumont, S., Smith, S. B., Tinoco, I. Jr & Bustamante, C. Equilibrium information from nonequilibrium measurements in an experimental test of the Jarzynski equality. *Science* **296**, 1832–1835 (2002).
- Ritort, F., Bustamante, C. & Tinoco, I. Jr A two-state kinetic model for the unfolding of single molecules by mechanical force. *Proc. Natl Acad. Sci. USA* **99**, 13544–13548 (2002).
- Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free-energy differences. *Phys. Rev. E* **60**, 2721–2726 (1999).
- Smith, S. B., Cui, Y. & Bustamante, C. An optical-trap force transducer that operates by direct measurement of light momentum. *Methods Enzymol.* **361**, 134–162 (2003).
- McManus, M. T., Petersen, C. P., Haines, B. B., Chen, J. & Sharp, A. P. Gene silencing using micro-RNA designed hairpins. *RNA* **8**, 842–850 (2002).
- Serganov, A. et al. Role of conserved nucleotides in building the 16S rRNA binding site for ribosomal protein S15. *J. Mol. Biol.* **305**, 785–803 (2002).
- Ritort, F. Work fluctuations, transient violations of the second law and free-energy recovery methods. *Semin. Poincaré* **2**, 193–226 (2003).
- Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* **78**, 2690–2693 (1997).
- Park, S. & Schulten, K. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* **120**, 5946–5961 (2004).
- Liphardt, J., Onoa, B., Smith, S. B., Tinoco, I. Jr & Bustamante, C. Reversible unfolding of single RNA molecules by mechanical force. *Science* **292**, 733–737 (2001).
- Zuckerman, D. M. & Woolf, T. B. Theory of systematic computational error in free energy differences. *Phys. Rev. Lett.* **89**, 180602 (2002).
- Gore, J., Ritort, F. & Bustamante, C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proc. Natl Acad. Sci. USA* **100**, 12564–12569 (2003).
- Shirts, R., Bair, E., Hooker, G. & Pande, V. S. Equilibrium free energies from nonequilibrium measurements using maximum likelihood methods. *Phys. Rev. Lett.* **91**, 140601 (2003).
- Bennett, C. H. Efficient estimates of free energy differences from Monte Carlo data. *J. Comp. Phys.* **22**, 245–268 (1976).
- SantaLucia, J. Jr & Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 415–440 (2004).
- Robertus, D. J. et al. Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* **250**, 546–551 (1974).
- Long, D. M. & Uhlenbeck, O. C. Self-cleaving catalytic RNA. *FASEB J.* **7**, 25–30 (1993).
- Cate, J. H. & Doudna, J. A. Metal binding sites in the major groove of a large ribozyme domain. *Structure* **4**, 1221–1229 (1996).
- Zuker, M. Mfold web server for nucleic acid folding and hybridization predictions. *Nucleic Acids Res.* **31**, 3406–3415 (2003).
- Turner, D. H. in *Nucleic Acids: Structures, Properties and Functions* (eds Bloomfield, D. A., Crothers, D. M. & Tinoco, I. Jr) Ch. 7 (Univ. Press, New York, 2000).
- Carrion-Vazquez, M., et al. Protein nanomechanics studied by AFM single-molecule force spectroscopy. In *Emerging Techniques in Biophysics* (eds Arrondo, J. L. R. & Alonso, A.) (Biophys. Monogr. Ser., Springer, Heidelberg, in the press).
- Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).
- Manosas, M. & Ritort, F. Thermodynamic and kinetic aspects of RNA pulling experiments. *Biophys. J.* **88**, 3224–3242 (2005).
- Hummer, G. Fast-growth thermodynamics integration: Error and efficiency analysis. *J. Chem. Phys.* **114**, 7330–7337 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Hummer and A. Szabo for many discussions and G. E. Crooks, D. Chandler and J. Liphardt for a careful reading of the manuscript. F.R. was supported by the Spanish Research council and the Catalan Government (Distinció de la Generalitat). C.J. was supported by an NIH grant and the US Department of Energy. I.T. was supported by an NIH grant. C.B. was supported by the Howard Hughes Medical Institute and the David and Lucile Packard Foundation.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.B. (carlos@alice.berkeley.edu) or F.R. (ritort@ffn.ub.es).

Ionic colloidal crystals of oppositely charged particles

Mirjam E. Leunissen^{1*}, Christina G. Christova^{1*}, Antti-Pekka Hynninen¹, C. Patrick Royall^{1†}, Andrew I. Campbell¹, Arnout Imhof¹, Marjolein Dijkstra¹, René van Roij² & Alfons van Blaaderen¹

Colloidal suspensions are widely used to study processes such as melting, freezing^{1–3} and glass transitions^{4,5}. This is because they display the same phase behaviour as atoms or molecules, with the nano- to micrometre size of the colloidal particles making it possible to observe them directly in real space^{3,4}. Another attractive feature is that different types of colloidal interactions, such as long-range repulsive^{1,3}, short-range attractive⁵, hard-sphere-like^{2–4} and dipolar³, can be realized and give rise to equilibrium phases. However, spherically symmetric, long-range attractions (that is, ionic interactions) have so far always resulted in irreversible colloidal aggregation⁶. Here we show that the electrostatic interaction between oppositely charged particles can be tuned such that large ionic colloidal crystals form readily, with our theory and simulations confirming the stability of these structures. We find that in contrast to atomic systems, the stoichiometry of our colloidal crystals is not dictated by charge neutrality; this allows us to obtain a remarkable diversity of new binary structures. An external electric field melts the crystals, confirming that the constituent particles are indeed oppositely charged. Colloidal model systems can thus be used to study the phase behaviour of ionic species. We also expect that our approach to controlling opposite-charge interactions will facilitate the production of binary crystals of micrometre-sized particles, which could find use as advanced materials for photonic applications⁷.

We recently developed a versatile colloidal model system of charged, sterically stabilized, polymethylmethacrylate (PMMA) spheres in a nearly density-matched mixture of cyclohexyl bromide (CHB) and cis-decalin^{3,8}. We simultaneously matched the refractive indices to eliminate the van der Waals forces. The range of the electrostatic interactions can be tuned with a salt like tetrabutylammonium bromide (TBAB) and is characterized by the Debye screening length, κ^{-1} , which for a number density $2c$ of monovalent cations and anions is given by

$$\kappa^{-1} = (8\pi\lambda_B c)^{-\frac{1}{2}} \quad (1)$$

Here the Bjerrum length is $\lambda_B = e^2/(4\pi\epsilon\epsilon_0 k_B T) = 10.0$ nm, with ϵ the relative dielectric constant of our solvent, ϵ_0 the dielectric permittivity of vacuum, e the elementary charge, k_B Boltzmann's constant and T the absolute temperature. From measurements of the three-dimensional radial distribution function we found that these suspensions are well-described by a pair-wise screened Coulomb potential $V_{ij}(r)$ (ref. 8):

$$\frac{V_{ij}(r)}{k_B T} = Z_i Z_j \lambda_B \frac{e^{\kappa(a_i + a_j)}}{(1 + \kappa a_i)(1 + \kappa a_j)} \frac{e^{-\kappa r}}{r} \quad (2)$$

where r is the distance between two particles with respective radii a_{ij} and charges $Z_{ij}e$.

Here we increase the complexity by investigating suspensions that consist of two types of colloids with opposite, dissimilar charges Z_1 and Z_2 and different radii a_1 and a_2 . For binary systems, much less work is reported than for one-component suspensions and it mostly deals only with long-range repulsive⁹ and hard-sphere-like interactions^{10,11} (although several researchers recently observed binary crystals of nanometre-sized colloids that probably formed in the presence of attractions^{12,13}).

We discovered that in our apolar suspensions the addition of TBAB salt offers control over the particle charge, besides regulating the screening length. The charge even reverses from plus to minus at moderate salt concentrations⁸. Given the many components involved in the synthesis of the colloids¹⁴, it is not surprising that we obtain a slightly different charge reversal point for each batch of particles. Exploiting this, we prepared salt-containing mixtures of differently labelled fluorescent spheres, thereby establishing binary systems of colloids that carry small, opposite charges. Below, we present our observations on same-size spheres, followed by mixtures of large and small spheres.

For a relatively broad range of parameters we observed complete freezing into large, caesium-chloride-type (CsCl) single crystals, that consist of several tens of particle layers measuring up to $300 \times 300 \mu\text{m}$ (Supplementary Fig. 1). This even happens for colloids as large as the ones in Fig. 1a–d. These plus and minus particles had radii $a_1 = 1.08$ and $a_2 = 0.99 \mu\text{m}$ respectively, that is, a size ratio a_2/a_1 close to unity (0.92). Typical suspensions with a 1:1 particle number ratio had an overall volume fraction $\phi \approx 0.12$ and contained $\sim 60 \mu\text{M}$ TBAB, corresponding to $\kappa^{-1} = 285$ nm. From electrophoretic mobility measurements we found approximate charges of $Z_1 = +110$ and $Z_2 = -75$ (± 10 , s.d.). As expected, for significantly higher charges only irreversible aggregation occurred. We stress that we did the measurements both with dilute suspensions of the individual particle species and *in situ* in the concentrated binary samples used for phase behaviour studies (see Methods section for details).

The CsCl character, that is, two interlaced simple cubic lattices together forming a body-centred cubic (b.c.c) lattice, was confirmed by quantitative confocal microscopy and by measuring the Bragg angles with light scattering (Supplementary Methods, Supplementary Table 1 and Supplementary Fig. 2). We found $2.36 \mu\text{m}$ for the lattice parameter, which translates into a packing fraction of 0.71. Within the experimental error, this is the closest CsCl-type packing possible for these spheres.

We followed the nucleation and growth of the crystals over time (Supplementary Fig. 3). Within ~ 35 h we see the first crystallites, their random orientation indicating homogeneous nucleation. This nucleation time is remarkably short for binary colloids, because slow

¹Soft Condensed Matter, Debye Institute, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands. ²Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CE Utrecht, The Netherlands. [†]Present address: Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 1538505, Japan.

*These authors contributed equally to this work.

diffusion and fast sedimentation were expected to hinder crystal growth for our 2- μm spheres. Nevertheless, without fine-tuning the conditions, we observed significantly faster crystallization than, for instance, the four days reported for certain optimized hard-sphere systems of much smaller particles^{10,11}. This ease of growth is important for the production of advanced materials, like photonic crystals, because bandgaps in the near-infrared require a dense, regular packing of micrometre-sized spheres⁷.

At $\varphi = 0.30$ and 190 μM TBAB ($\kappa^{-1} = 195$ nm) we observed a close-packed lattice with the plus and minus particles randomly distributed over the sites (Fig. 1e). This so-called 'solid solution' had a lattice spacing $\sim 6\%$ larger than the densest possible packing (packing fraction 0.74). Semi-quantitative observations indicate that this different structure, which was recently predicted in simulations

of the restricted primitive model¹⁵, is due to lower particle charges.

We also looked at binary crystals of two different materials, because the ability to create these is important for applications. We used PMMA ($a_1 = 0.58$ μm , positive) and silica ($a_2 = 0.52$ μm , negative) particles in a 53 vol.%/47 vol.% CHB-decalin mixture ($\varphi \approx 0.13$ – 0.20 and ~ 160 μM TBAB). Despite the large density contrast (~ 1.2 g cm^{-3} for PMMA versus 2.0 g cm^{-3} for silica) the electrostatic attractions are strong enough to overcome differences in sedimentation. This produced quick growth of CsCl-type crystals again (Fig. 1f, g).

Previously, we demonstrated that PMMA can be burned away from composite crystals, leaving the silica spheres behind¹⁶. Using this procedure, one could turn a CsCl-type crystal into a simple cubic lattice, a structure that has not yet been realized by self-organization of colloids. Moreover, silica-only structures are important for the production of photonic crystals with 'inverse' index contrast, as this often requires a high temperature reaction⁷.

We confirmed that the particles inside the binary crystals are still oppositely charged by applying an external, static electric field. Figure 2a shows how a CsCl-type crystal melts as the differently labelled particles move towards opposite electrodes. This also demonstrates

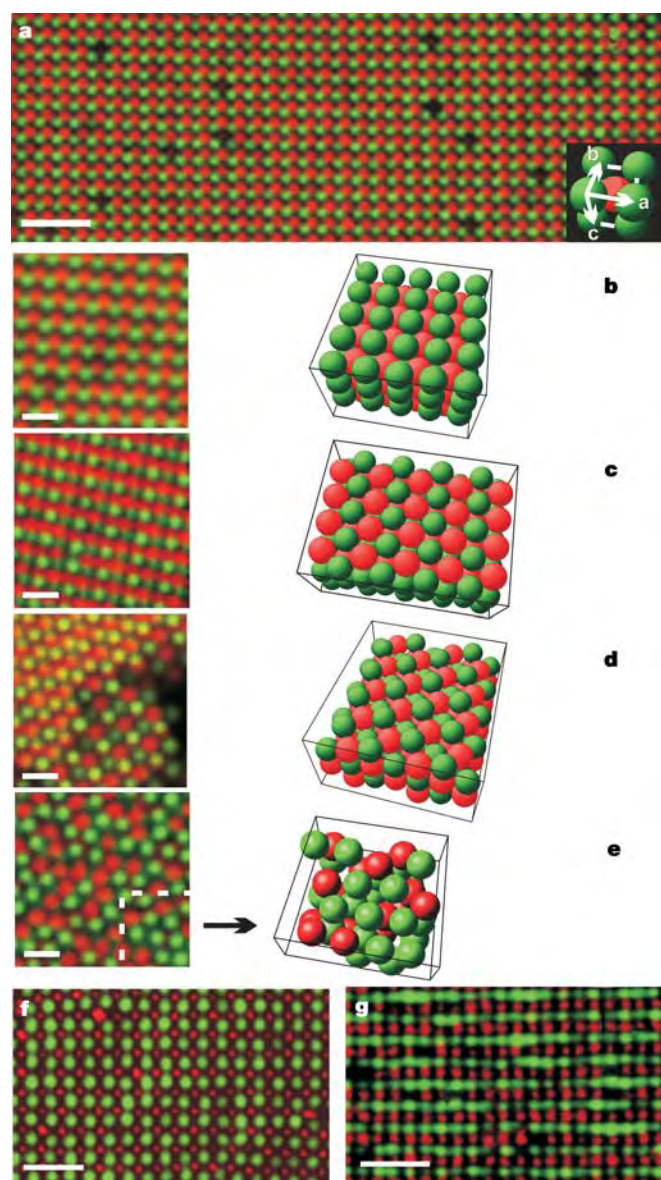


Figure 1 | CsCl-type binary crystals. **a–e**, Positive (red, radius 1.08 μm) and negative (green, 0.99 μm) PMMA-spheres. **a**, Confocal micrograph of a large (100) plane (scale bar, 10 μm). Inset, the cubic CsCl-type unit cell. **b–d**, Close-up of the (100), (110) and (111) planes plus corresponding models. **e**, 'Solid solution'. The stacking of the hexagonal layers is visible in the box with rendered coordinates. The model spheres have a smaller radius for clarity. **f**, CsCl (100) and **g**, (110) planes with positive PMMA (green, 0.52 μm) and negative silica (red, 0.58 μm) particles. Scale bars in **b–g**, 4 μm . All particles were dispersed in TBAB-containing CHB-decalin.

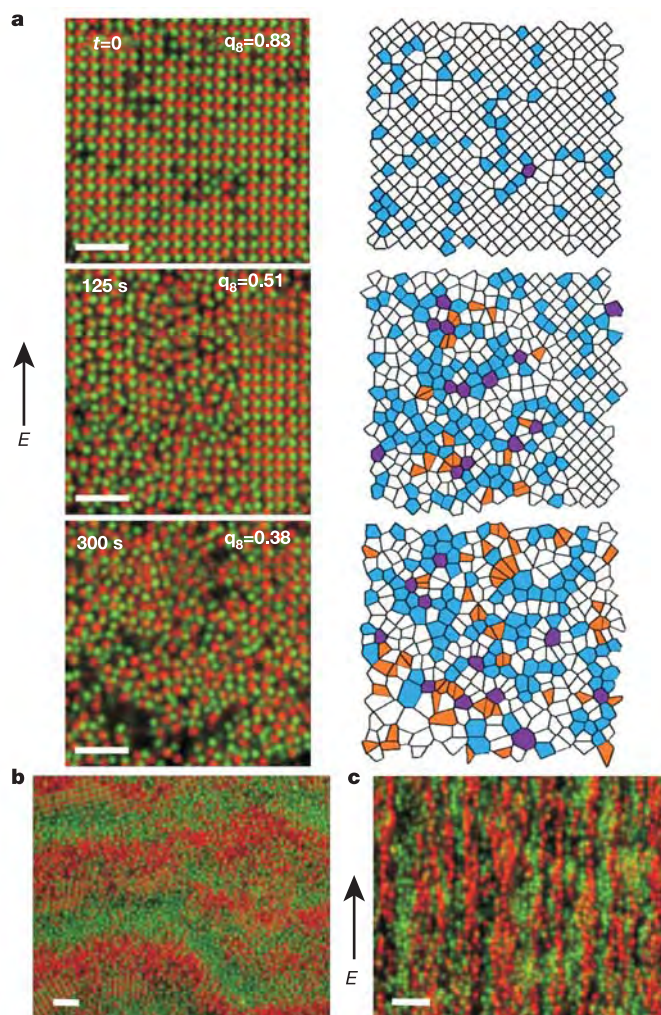


Figure 2 | Electric-field-induced melting. **a**, Confocal snapshots of the melting process induced by a static electric field ($E = 7$ V mm^{-1}) at $t = 0$, 125 and 300 s. The frame-averaged value of the bond order parameter q_8 is indicated. In the Voronoi constructions orange polygons are threefold-coordinated; white, fourfold; blue, fivefold; purple, sixfold³¹. **b**, Jammed 'band' formation in a dense sample. **c**, Dynamic lane formation at lower density and 80 V mm^{-1} . All scale bars are 10 μm .

that electric fields are a powerful way to manipulate these soft colloidal crystals. Here, a mere 7 V mm^{-1} sufficed. Furthermore, we note that this system is very similar to a certain electronic ink ('e-ink')¹⁷, although e-ink aggregates.

In some cases the electric field induced interesting pattern formation. Dense suspensions ($\phi > 0.40$) in fairly high fields ($\sim 20 \text{ V mm}^{-1}$) form 'bands' perpendicular to the field direction (Fig. 2b). The different particles initially move in opposite directions, but soon form a glassy state, which no longer shows any diffusion or field-induced drift. This process is reversible: the system 'un-jams' when the field is switched off. In relatively dilute suspensions ($\phi \approx 0.20\text{--}0.30$) the colloids remain mobile and another pattern appears, with the oppositely driven particles segregated into 'lanes' along the field direction (Fig. 2c). This so-called 'lane formation' is a first-order out-of-equilibrium phase transition^{18,19}. It can now be studied in a controlled way and quantitatively at the single particle level.

We also investigated mixtures of large (L) and small (S) spheres, with a considerable size difference ($a_2/a_1 = 0.31$). Using a 1:8 (L:S number density) suspension of negative ($a_1 = 1.16 \mu\text{m}$) and positive ($a_2 = 0.36 \mu\text{m}$) PMMA-particles ($\phi \approx 0.11$, estimated TBAB-concentration $\sim 120 \mu\text{M}$) we observed crystals with LS_6 stoichiometry (Fig. 3). This new structure has a face-centred orthorhombic lattice of the large spheres. One small sphere occupies each of the tetrahedral holes, while four small spheres are found in each of the octahedral holes (two slightly above and two slightly below each ab -plane of large particles in Fig. 3a and d). The remaining small spheres are expelled from the crystal. (After submission of this Letter Pusey *et al.* drew our attention to their observations of a slightly different colloidal LS_6 -structure, which has a b.c.c. symmetry²⁰. However, this was reported to be a binary hard-sphere mixture and no further details were given. We speculate that these binary crystals were also grown from oppositely charged colloids). Interestingly, these colloidal structures resemble the crystals that are formed by

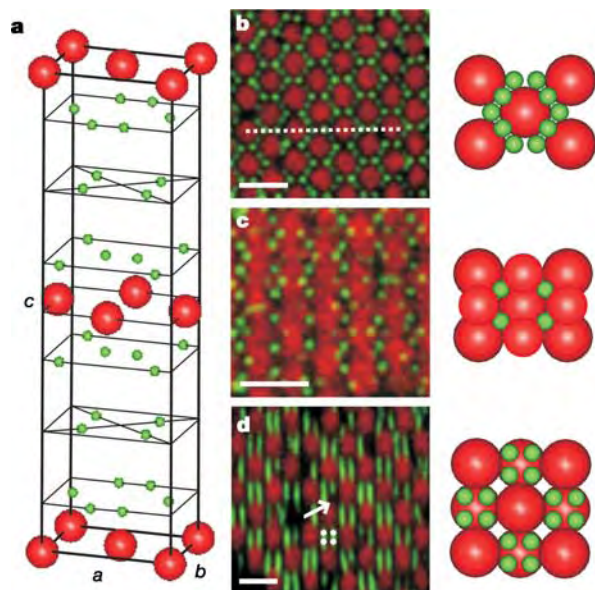


Figure 3 | LS_6 -type binary crystals. **a–d**, Positive (green, radius $0.36 \mu\text{m}$) and negative (red, $1.16 \mu\text{m}$) PMMA-particles in TBAB-containing CHB-decalin, forming a structure with LS_6 stoichiometry. **a**, Unit cell with observed lattice spacings $a = 4.00$, $b = 2.84$ and $c = 4.7 \mu\text{m}$ (not to scale for clarity). **b**, **c**, Confocal images and models of different ab -projections, showing a layer of large and several layers of small particles (**b**) and a plane with only small particles (**c**). **d**, ac -cut along the line in **b**. As the microscope could not completely resolve the four small particles in each octahedral hole, we indicated their positions with dots. The arrow indicates a missing particle. All scale bars are $4 \mu\text{m}$.

certain alkali-metal intercalation compounds of the fullerene C_{60} with the same stoichiometry²¹. In these compounds there is a large size difference between the constituent ions too.

At slightly lower ionic strength we found another new structure, LS_8 , for which no atomic or molecular counterpart has yet been identified. In this case, each large colloid is surrounded by six small spheres that occupy the trigonal interstices of the hexagonal planes. Above and below this 'mixed' layer are planes of small particles only in a Kagomé-type arrangement. This three-plane unit repeats itself in such a way that the large colloids form an $ABAB$ -stacking.

It also proved possible to establish conditions under which the small spheres were essentially uncharged, while the large spheres were slightly charged, as judged from their electrophoretic mobility and the lattice spacing in crystals of the individual particle species. In a L:S = 1:4 suspension with $\phi \approx 0.23$ and no added salt

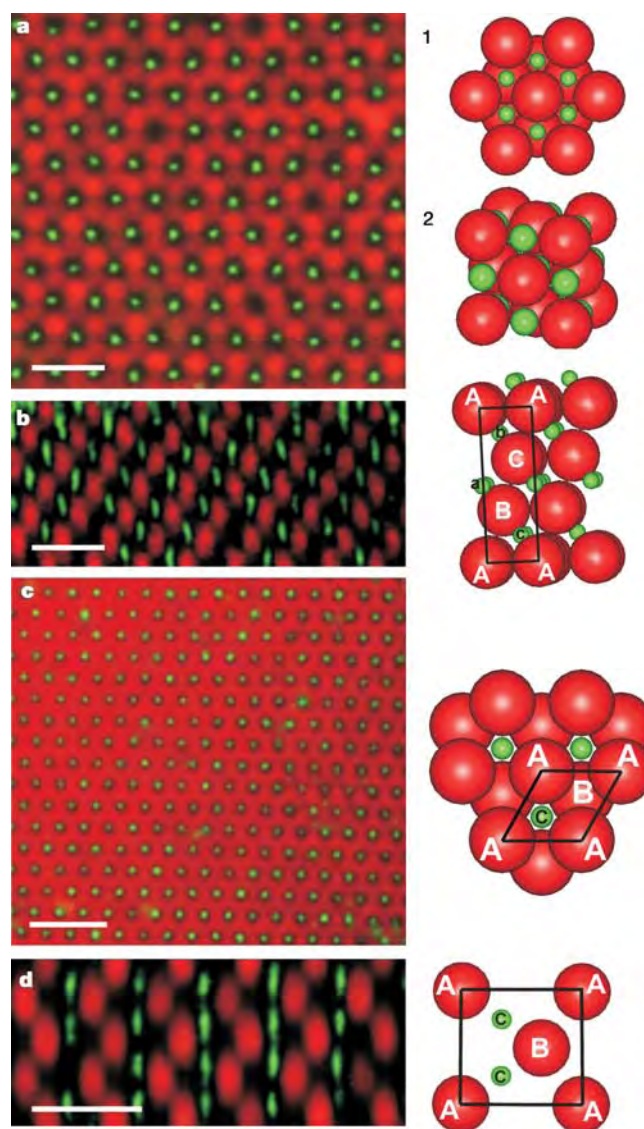


Figure 4 | LS -type binary crystals. Charged (red, radius $1.16 \mu\text{m}$) and uncharged (green, $0.36 \mu\text{m}$) PMMA-particles in CHB-decalin. **a**, **b**, NaCl-type crystal. **a**, Confocal image of the hexagonal plane and the unit cell in a hexagonal (1) and a cubic (2) representation (small spheres enlarged for clarity). **b**, Plane perpendicular to the hexagonal close packed layers, showing the ABC -stacking of both the large and small particles. **c**, **d**, NiAs-type crystal. **c**, Superposition of confocal images of ten layers and the corresponding model. **d**, Plane as in **b**, the model shows the ABA -stacking of the large and the ccc -stacking of the small particles (spheres not to scale). All scale bars are $8 \mu\text{m}$.

($\kappa^{-1} = 1.57 \mu\text{m}$) we observed coexistence of two different crystal structures, both with LS stoichiometry: sodium chloride type (NaCl) and nickel arsenide type (NiAs) (Fig. 4). The latter structure has a hexagonal-close-packed (h.c.p.) stacking of the large spheres with all the octahedral sites occupied by the small spheres, which form a simple hexagonal sublattice. In the NaCl-type structure the large spheres form an face-centred cubic (f.c.c.) lattice. The small spheres fill the octahedral holes again, but now as an interpenetrating f.c.c. lattice. Recently, LS-type structures were also observed in hard-sphere suspensions¹¹ and in systems of nanocrystals with attractions¹².

Although ionic crystals have been theoretically investigated in many papers, we found only one²² that deals with oppositely charged spheres interacting through a screened Coulomb potential. (After submitting this Letter we became aware of a thesis²³ that contains

theoretical investigations and computer simulations of the interactions between oppositely charged colloids and experimental attempts to realise binary crystals of such particles.) Unfortunately, it restricts itself to a single set of interaction parameters, so we explored the stability of different crystal structures further ourselves. We did this for the size ratios studied experimentally (~ 1 and 0.31), as a function of the charge ratio $Q = -Z_1/Z_2$ and the reduced screening constant κa_1 .

For long screening lengths (equation (1)) our oppositely charged particles resemble ionic systems, but with differences. Most importantly, colloids together with their surrounding 'diffuse layer' of counter-ions are charge-neutral objects. Therefore, the stoichiometry is not dictated by charge neutrality: even with a considerable charge asymmetry colloids can form crystals with LS (or some other) stoichiometry, as in Fig. 1.

The equilibrium structure of a crystal is determined by geometric requirements, stemming from short-range repulsions, together with the potential energy of the lattice and the entropy associated with thermal effects. For ionic systems, the potential energy strongly depends on the Coulomb interactions, which we describe by the pair-potential of equation (2). This probably is a good approximation for the like-charge repulsions⁸, but its accuracy for the plus-minus attractions is not well known (see below). We further assume a vanishingly small osmotic pressure, such that the crystals are self-supported by their cohesive energy and coexist with a zero-density vapour.

The phase diagrams in Fig. 5 display structures with minimal Madelung energy (potential energy per particle in the perfect crystal; we assume touching neighbours), in which we neglect entropy. In this zero-temperature limit the phase diagram depends on the charge ratio and not on the absolute contact energies V_{ij} (equation (2) evaluated at $r = a_i + a_j$). The absolute charges only become relevant in the Monte Carlo simulations that we performed in order to assess the importance of entropy at ambient temperature (see Methods section).

The experimentally determined charges (of the particles in Fig. 1a–d) give a contact energy V_{12} of $\sim -2k_B T$ for the plus-minus interactions. However, we bear in mind that the electric field gradient becomes very steep near contact. It is not known whether the Poisson–Boltzmann approach underlying equation (2) is valid in this case, nor what is the correct boundary condition (that is, constant surface charge versus potential or something in between). Therefore, we chose several contact energies and examined how this affected the stability lines.

Figure 5a shows that the NaCl and CsCl structures dominate for same-size colloids. If we include entropic effects, both the NaCl–CsCl phase boundary and the NaCl melting line shift to higher κa_1 (more screening). In our experiments, the clear particle motion around the lattice positions indicates that entropy is indeed important. The observed CsCl-type crystals (Fig. 1a–d) lie inside the predicted CsCl regime. Their Madelung energy was calculated to be $\sim -6k_B T$ (per particle). In the 'molten' regime (positive NaCl Madelung energy) we find many other crystal phases, but this will be discussed elsewhere.

The phase diagram for size ratio 0.31 (Fig. 5b) nicely demonstrates how the stoichiometry follows the charge ratio at low κa_1 . We obtained this diagram with the common tangent construction, taking the dilute phase rich in small colloids. Of course, this is just one of many possible 'slices' through the multidimensional phase diagram. If we choose the composition of the dilute phase differently, we obtain a different diagram. However, the main features, such as the order of the phases along the Q -axis, remain the same. Thus, the charge ratio in the experimentally observed LS_6 -structure is probably smaller than six. The diagram also contains a region where the LS_8 -structure is stable. The insets in Fig. 5b show a model of this structure; we will present experimental data elsewhere.

Our present experiments and those of others demonstrate that

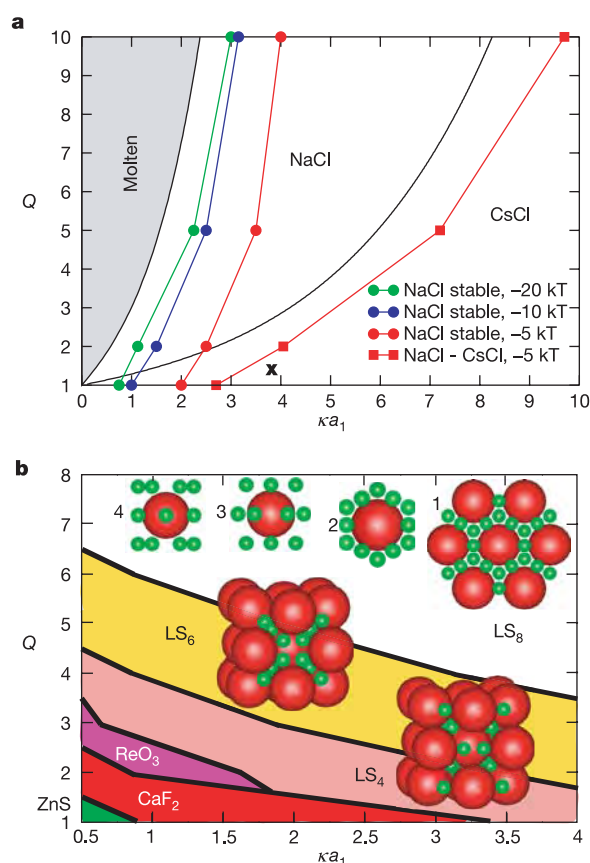


Figure 5 | Theoretical phase diagrams. Based on screened-Coulomb pair-interactions (equation (2)) for different screening constant, κa_1 , and charge ratio $Q = -Z_1/Z_2$ at a size ratio a_2/a_1 of 1 (a) and 0.31 (b). Full curves indicate crystal–crystal transitions from Madelung energy calculations (zero-temperature; black) and finite-temperature, zero-pressure MC-simulations with a contact energy V_{12} of $-20k_B T$ (green), $-10k_B T$ (blue) and $-5k_B T$ (red). a, The NaCl-structure melting line and NaCl–CsCl phase boundary. The cross indicates the estimated state point of the experimentally observed CsCl crystals (Fig. 1a–d). NiAs, CuAu and ZnS structures, which we also evaluated, are un- or metastable. b, Crystal phases coexisting with a dilute gas of small colloids, from Madelung energies. Unlike LS_6 and LS_8 the LS_4 structure was not experimentally observed, but generated by removing two small particles from the octahedral holes of LS_6 (compare the small colloid configuration around the central large sphere in the two models). In the LS_8 structure a unit of three particle planes (insets 3, 4; not to scale) repeats itself to form a hexagonal ABAB-stacking of the large colloids. Insets: 1, projection of the three basic planes of LS_8 , 2, configuration around a large colloid, 3, side-view, 4, side-view rotated by 90 degrees. We also evaluated structures of NaCl, CsCl, NiAs, CuAu, AlB₂, Cu₃Au, Al₃Ti and CaCu₅, but these were found to be unstable.

equilibrium phase behaviour of oppositely charged colloids occurs for a fairly broad range of parameters and we expect that binary crystals should be observable for particle sizes covering the entire colloidal range. Evidence of crystallizing plus and minus spheres has been observed under conditions different from those used here²⁴. We also suspect that Underwood *et al.*²⁵ may have ruled out the presence of opposite charges too quickly as the origin of the attractions that caused their CsCl-type crystals. All of these studies involved organic solvents, but we expect that ionic crystals can also be obtained in more polar solvents, such as water. Although it is more difficult to keep the van der Waals forces sufficiently small in water, this was recently achieved for a suspension with attractive interactions mediated through complementary strands of DNA on the particles²⁶. A mixture with different complementary sequences for different particles would actually mimic our system for short screening lengths.

METHODS

Particle synthesis. We used polymethylmethacrylate (PMMA) and silica particles. The former were made by dispersion polymerization, covalently labelled with the fluorophore 7-nitrobenzo-2-oxa-1,3-diazol (NBD) or rhodamine isothiocyanate (RITC) and sterically stabilized with poly-12-hydroxystearic acid (PHSA)¹⁴. These particles had a size polydispersity of 3–5% (from light scattering and electron microscopy). We prepared the silica colloids using a modified Stöber synthesis²⁷, which includes continuous-feed seeded growth, resulting in an RITC-labelled core of radius 200 nm surrounded by an unlabelled shell. The particles were then coated with 3-methacryloxypropyltrimethoxysilane (TPM) and stabilized with PHSA. The total radius was 0.58 μm , with 4% polydispersity.

Sample preparation. Unless indicated otherwise, the colloids were dispersed in a mixture of cyclohexyl bromide (CHB, Fluka) and 27.2 wt% *cis*-decalin (Sigma-Aldrich) that nearly matches the density and refractive index of PMMA. We purified the CHB as described before⁸, but skipped the distillation. The particle charge and the range of the electrostatic interactions were tuned by adding tetrabutylammonium bromide salt (TBAB, Sigma-Aldrich)⁸. We let separate suspensions of the individual particle species equilibrate for several hours before mixing them together. We then confined the binary mixture to glass capillaries (VitroCom) and studied it by means of confocal laser scanning microscopy, extracting the three-dimensional particle coordinates as described before⁸. Gradient samples served to explore the phase behaviour as a function of the salt concentration. We made these by filling a capillary partly with a salt-free suspension and partly with a TBAB solution and allowing it to form macroscopic gradients in a few days' time (at constant particle volume fraction).

Electrical conductivity measurements. We estimated the Debye screening length of our suspensions by measuring the conductivity of the particle-free solvent mixtures (with a Scientifica 627 conductivity meter) and then applying Walden's rule⁸. We used literature values for the limiting equivalent conductance of TBAB in water and a viscosity of 2.217 centipoise at 25 °C (measured with a Schott ViscoSystem) for the solvent mixture.

Electrophoretic mobility measurements. Particle charges were quantified by means of electrophoresis (using a Coulter Delsa 440SX) on dilute suspensions (volume fraction 0.0015) in the CHB-decalin mixture⁸. The run parameters were: 25 V, 2.0 s on/0.5 s off, and a total run time of 60 s. We identified the stationary layers by Komagata linearization²⁸ and related the electrophoretic mobility to the zeta potential via the O'Brien and White scheme²⁹. From dielectric spectroscopy (with an HP 4294A impedance analyser) the dielectric constant of the solvent mixture was found to be 5.59.

We also estimated the charges in the denser microscopy samples. The sample cell consisted of a 0.1×1 mm capillary with two 50- μm -diameter nickel-alloy wires (Goodfellow) threaded through along the side walls. We applied a direct-current field (up to 25 V mm^{-1}) and captured the electrophoretic motion of individual particles shortly after they left the binary crystals (scan speed 1–3 frames/second). We determined the particles' mobility from the displacements between subsequent frames.

Theory and computer simulations. We parameterized our Madelung energy calculations and computer simulations by choosing values for the reduced screening constant κa_1 , the charge ratio $Q = -Z_1/Z_2$ and the contact energy $V_{12}(a_1 + a_2)$. These parameters, together with the size ratio a_2/a_1 , fully determine the pair potential of equation (2). We performed so-called 'constant-NPT' Monte Carlo simulations, thus keeping the number of particles N , the pressure P and the temperature T fixed. If a crystal coexists with a low-density gas one can estimate its coexistence density with a simulation at zero-

pressure³⁰. In this way we obtained the NaCl-structure melting lines and the stabilities of all other crystal structures. The NaCl-CsCl phase line at a contact energy V_{12} of $-5k_B T$ was determined by first minimizing the free energy of both phases with respect to density and then finding the point where the free energy per particle of the two phases is equal. We calculated the free energies from 'constant-NVT' Monte Carlo simulations (fixing the number of particles N , the volume V and the temperature T) using the Frenkel-Ladd method³⁰. We did all the simulations in a cubic box with periodic boundary conditions, taking 216 and 250 particles for the NaCl- and CsCl-crystals respectively.

We determined the Madelung energy with the potential energy calculation in the Monte Carlo code, assuming contact between neighbouring spheres in a structure. The absolute value of V_{12} does not play a role as long as the same value is used throughout. For the LS_4 , LS_6 and LS_8 structures the energy minimum was found by varying their lattice parameters. We then drew up the phase diagram with the common tangent construction, taking the dilute phase rich in small colloids. All the other structures that we evaluated can be found in the legend of Fig. 5.

Received 11 April; accepted 20 June 2005.

- Hachisu, S., Kobayashi, Y. & Kose, A. Phase separation in monodisperse latexes. *J. Colloid Interf. Sci.* **42**, 342–348 (1973).
- Pusey, P. N. & van Megen, W. Phase behaviour of concentrated suspensions of nearly hard colloidal spheres. *Nature* **320**, 340–342 (1986).
- Yethiraj, A. & van Blaaderen, A. A colloidal model system with an interaction tunable from hard sphere to soft and dipolar. *Nature* **421**, 513–517 (2003).
- Kegel, W. K. & van Blaaderen, A. Direct observation of dynamical heterogeneities in colloidal hard-sphere suspensions. *Science* **287**, 290–293 (2000).
- Pham, K. N. *et al.* Multiple glassy states in a simple model system. *Science* **296**, 104–106 (2002).
- Islam, A. M., Chowdhry, B. Z. & Snowden, M. J. Heteroaggregation in colloidal dispersions. *Adv. Colloid Interf.* **62**, 109–136 (1995).
- Vlasov, Y. A., Bo, X. Z., Sturm, J. C. & Norris, D. J. On-chip natural assembly of silicon photonic bandgap crystals. *Nature* **414**, 289–293 (2001).
- Royall, C. P., Leunissen, M. E. & van Blaaderen, A. A new colloidal model system to study long-range interactions quantitatively in real space. *J. Phys. Condens. Matter* **15**, S3581–S3596 (2003).
- Hachisu, S. & Yoshimura, S. in *Physics of Complex and Supermolecular Fluids* (eds Safran, S. A. & Clark, N. A.) 221–240 (Wiley, New York, 1987).
- Schofield, A. B. Binary hard-sphere crystals with the cesium chloride structure. *Phys. Rev. E* **64**, 051403 (2001).
- Hunt, N., Jardine, R. & Bartlett, P. Superlattice formation in mixtures of hard-sphere colloids. *Phys. Rev. E* **62**, 900–913 (2000).
- Saunders, A. E. & Korgel, B. A. Observation of an AB phase in bidisperse nanocrystal superlattices. *Chemphyschem* **6**, 61–65 (2005).
- Redl, F. X., Cho, K. S., Murray, C. B. & O'Brien, S. Three-dimensional binary superlattices of magnetic nanocrystals and semiconductor quantum dots. *Nature* **423**, 968–971 (2003).
- Bosma, G. *et al.* Preparation of monodisperse, fluorescent PMMA-latex colloids by dispersion polymerization. *J. Colloid Interf. Sci.* **245**, 292–300 (2002).
- Bresme, F., Vega, C. & Abascal, J. L. F. Order-disorder transition in the solid phase of a charged hard sphere model. *Phys. Rev. Lett.* **85**, 3217–3220 (2000).
- Velikov, K. P., Christova, C. G., Dullens, R. P. A. & van Blaaderen, A. Layer-by-layer growth of binary colloidal crystals. *Science* **296**, 106–109 (2002).
- Gelinck, G. H. *et al.* Flexible active-matrix displays and shift registers based on solution-processed organic transistors. *Nature Mater.* **3**, 106–110 (2004).
- Dzubiella, J., Hoffmann, G. P. & Löwen, H. Lane formation in colloidal mixtures driven by an external field. *Phys. Rev. E* **65**, 021402 (2002).
- Löwen, H. & Dzubiella, J. Nonequilibrium pattern formation in strongly interacting driven colloids. *Faraday Discuss.* **123**, 99–105 (2003).
- Pusey, P. N. General discussion. *Faraday Discuss.* **123**, 177 (2003).
- Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. *Science of Fullerenes and Carbon Nanotubes* Ch. 8 (Academic, London, 1996).
- Caballero, J. B., Puertas, A. M., Fernandez-Barbero, A. & de las Nieves, F. J. Oppositely charged colloidal binary mixtures: A colloidal analog of the restricted primitive model. *J. Chem. Phys.* **121**, 2428–2435 (2004).
- Maskaly, G. R. Attractive Electrostatic Self-Assembly of Ordered and Disordered Heterogeneous Colloids. PhD thesis, Massachusetts Institute of Technology (2005).
- Bartlett, P. & Campbell, A. I. Three-dimensional binary superlattices of oppositely charged colloids. *Phys. Rev. Lett.* (in the press).
- Underwood, S. M., van Megen, W. & Pusey, P. N. Observation of colloidal crystals with the cesium-chloride structure. *Physica A* **221**, 438–444 (1995).
- Biancaniello, P. L., Kim, A. J. & Crocker, J. C. Colloidal interactions and self-assembly using DNA hybridization. *Phys. Rev. Lett.* **94**, 058302 (2005).
- van Blaaderen, A. & Vrij, A. Synthesis and characterization of colloidal dispersions of fluorescent, monodisperse silica spheres. *Langmuir* **8**, 2921–2931 (1992).

28. Hunter, R. J. *Zeta Potential in Colloid Science* Ch. 4 (Academic, London, 1981).
29. O'Brien, R. W. & White, L. R. Electrophoretic mobility of a spherical colloidal particle. *J. Chem. Soc. Faraday Trans. II* **74**, 1607–1626 (1978).
30. Frenkel, D. & Smit, B. *Understanding Molecular Simulations* Ch. 5, 10 (Academic, London, 2002).
31. Weiss, J. A., Oxtoby, D. W., Grier, D. G. & Murray, C. A. Martensitic transition in a confined colloidal suspension. *J. Chem. Phys.* **103**, 1180–1190 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. P. A. Dullens, D. Derks, N. A. M. Verhaegh, P. Vergeer and C. M. van Kats for particle synthesis, A. D. Hollingsworth for solvent characterization, J. H. J. Thijssen for help with the pictures of the Bragg reflections and both P.N. Pusey and A.B. Schofield for pointing out the resemblance between our LS_6 -structure and certain fullerene compounds. This

work is part of the research program of the 'Stichting voor Fundamenteel Onderzoek der Materie' (FOM), which is financially supported by the 'Nederlandse organisatie voor Wetenschappelijk Onderzoek' (NWO).

Author contributions M.E.L. and C.G.C. investigated the phase behaviour of the experimental binary systems, A.-P.H. and M.D. performed the computer simulations, A.-P.H. and R.v.R. calculated the Madelung energies, C.P.R. worked on the lane formation, A.I. on the Bragg scattering, A.I.C. made different plus/minus systems and A.v.B. initiated the work and co-wrote the paper together with M.E.L.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.E.L. (M.E.Leunissen@phys.uu.nl) or A.v.B. (A.vanBlaaderen@phys.uu.nl).

Orbital forcing of Cretaceous river discharge in tropical Africa and ocean response

Britta Beckmann¹, Sascha Flögel², Peter Hofmann³, Michael Schulz^{1,4} & Thomas Wagner⁵

The tropics have been suggested as the drivers of global ocean and atmosphere circulation and biogeochemical cycling during the extreme warmth of the Cretaceous period^{1,2}; but the links between orbital forcing, freshwater runoff and the biogeochemistry of continental margins in extreme greenhouse conditions are not fully understood. Here we present Cretaceous records of geochemical tracers for freshwater runoff obtained from a sediment core off the Ivory Coast that indicate that alternating periods of arid and humid African climate were driven by orbital precession. Our simulations of the precession-driven patterns of river discharge with a global climate model suggest that ocean anoxia and black shale sedimentation were directly caused by high river discharge, and occurred specifically when the northern equinox coincided with perihelion (the minimum distance between the Sun and the Earth). We conclude that, in a warm climate, the oceans off tropical continental margins respond rapidly and sensitively to even modest changes in river discharge.

Intervals of extreme warmth and enhanced sequestration of marine organic carbon (OC), termed oceanic anoxic events (OAEs), are one focus of current palaeoclimate research. They provide fundamental information on the functioning of biogeochemical cycles and their feedbacks during extreme conditions, especially when applied to areas expected to react sensitively to climate change, that is, continental margins and their associated sub-basins. Deposits from the eastern tropical Atlantic Ocean at ODP Site 959 off the Ivory Coast were obtained from such a setting (Fig. 1), providing a unique record of Coniacian to Campanian OC-rich sedimentation that covers the last of the Cretaceous OAEs (OAE 3). OAE 3 black shales have been reported from various basins surrounding the Atlantic and Tethyan Margin^{3–5}. Millennial-scale OAE 3 records from ODP Site 959 were used to develop a model for black shale formation in the tropical Atlantic⁶. Additionally, records of quartz and clay mineralogy document Upper Cretaceous atmospheric circulation that triggered latitudinal shifts of African climate belts. The records further imply that black shales formed in response to tropical continental discharge that induced stratification and reversals in surface ocean circulation within the deep Ivorian basin (DIB)⁶. Synchronous variations in trace metals⁷ and biomarkers of green sulphur bacteria⁸ further provide evidence for severe variations in redox conditions, with brief intervals of lower photic zone euxinia⁸. Here we present records that double the stratigraphic range of the previous profiles into the lower Campanian, allowing the deduction of long-term trends and shorter-scale variations in the tropical atmosphere–ocean system across one of the most important transitions in global climate over the past 150 Myr.

Coniacian–Campanian African climate variability is evident from fluctuations of K/Al and Ti/Al ratios (Fig. 2a, b). Al is mainly

confined to the fine-grained aluminosilicate fraction⁹, typically enriched in kaolinite, smectite and iron hydroxides¹⁰, and formed by intense chemical weathering under warm and humid conditions. K is associated with continental siliciclastics (that is, clay minerals and feldspar) that experienced only moderate amounts of chemical alteration¹¹. Ti is concentrated in heavy minerals and often transported via aeolian pathways into marine sediments¹². In agreement with vegetation simulations for the Campanian¹³, the source for Ti is assigned to desert areas of the proto-Kalahari in southern Africa, while K is associated with illite and K-feldspar derived from semi-arid tropical regions. Al from kaolinite and smectite is indicative of tropical weathering in equatorial northern Africa^{6,10,12}. The persistent periodicity of climate tracers at Site 959 documents that Upper Cretaceous climate was highly variable and modulated at different timescales. The average K/Al record exhibits a long-range increase towards the top of the profile (Fig. 2b). The progression in supply of detrital material from semi-arid regions supports a gradual aridification of Africa that started in the lower Santonian. The linearity of the complementary Ti/Al record indicates the existence of an arid proto-Kalahari remaining constant in extension and geographical position.

Time–frequency analyses of the K/Al and Ti/Al records support the existence of one dominant period with ~1.3 m wavelength, equivalent to ~28 kyr in nannofossil biozone CC15. Given uncertainties in

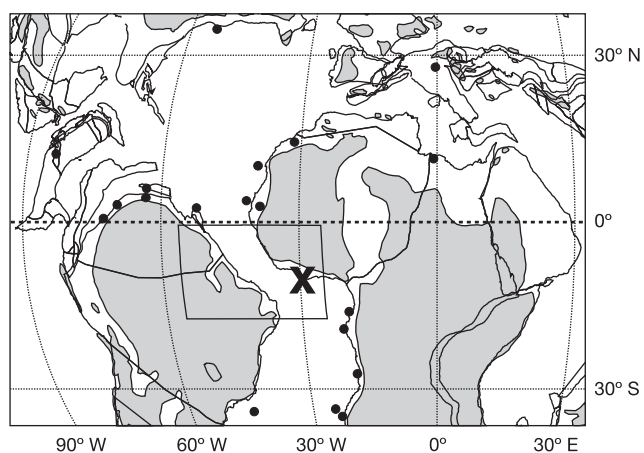


Figure 1 | Location of OAE 3 sites and investigated region for simulated river discharge. ODP Site 959 is marked by X; dots indicate other reported OAE 3 sites³; black box marks investigated region for simulated river discharge. All information is superimposed on the palaeogeographic map of the Upper Cretaceous ~80 Myr ago²⁰. Grey shaded areas represent land masses.

¹Department of Geosciences, University of Bremen, Klagenfurter Str., 28357 Bremen, Germany. ²IFM – GEOMAR Leibniz Institute of Marine Sciences, FB 1, Ocean Circulation and Climate Dynamics, Wischhofstr. 1–3, 24148 Kiel, Germany. ³Institute for Geology and Mineralogy, University of Cologne, Zùlpicher Str. 49a, 50674 Köln, Germany.

⁴Research Center Ocean Margins, University of Bremen, Leobener Str., 28357 Bremen, Germany. ⁵School of Civil Engineering and Geosciences, University of Newcastle, Newcastle upon Tyne NE1 7RU, UK.

Table 1 | Simulated river discharge

Orbital case	Perihelion at	Northern Hemisphere	Seasonal contrasts	Total annual discharge (km ³ yr ⁻¹)*	Wet season relative to annual discharge (%)	Wet season discharge, Mar.–Jun. (km ³ yr ⁻¹)	Dry season discharge, Jul.–Feb. (km ³ yr ⁻¹)	Thickness of annual freshwater lid (m)
Orb. A	Northern spring equinox	Warm spring/cold autumn	Max.	1,556 (+43%)	71	1,101 (+79%)	288 (–27%)	1.00
Orb. B	Northern winter solstice	Trans.	Trans.	1,365 (+25%)	67	917 (+49%)	276 (–30%)	0.88
Orb. C	Northern autumn equinox	Cold spring/warm autumn	Min.	1,091 (0%)	56	614 (0%)	393 (0%)	0.70
Orb. D	Northern summer solstice	Trans.	Trans.	1,343 (+23%)	63	851 (+39%)	407 (+4%)	0.86

Simulated continental freshwater discharge for the investigated region. The analysed area was situated in tropical, Upper Cretaceous Africa and South America (3,945,198.69 km²); the resulting thickness of the freshwater lid in the Equatorial Atlantic for all simulations is expressed as total values in km³ yr⁻¹, relative proportions, and metres. Max., maximum; trans., transitional; min., minimum.

* Relative to Orb. C.

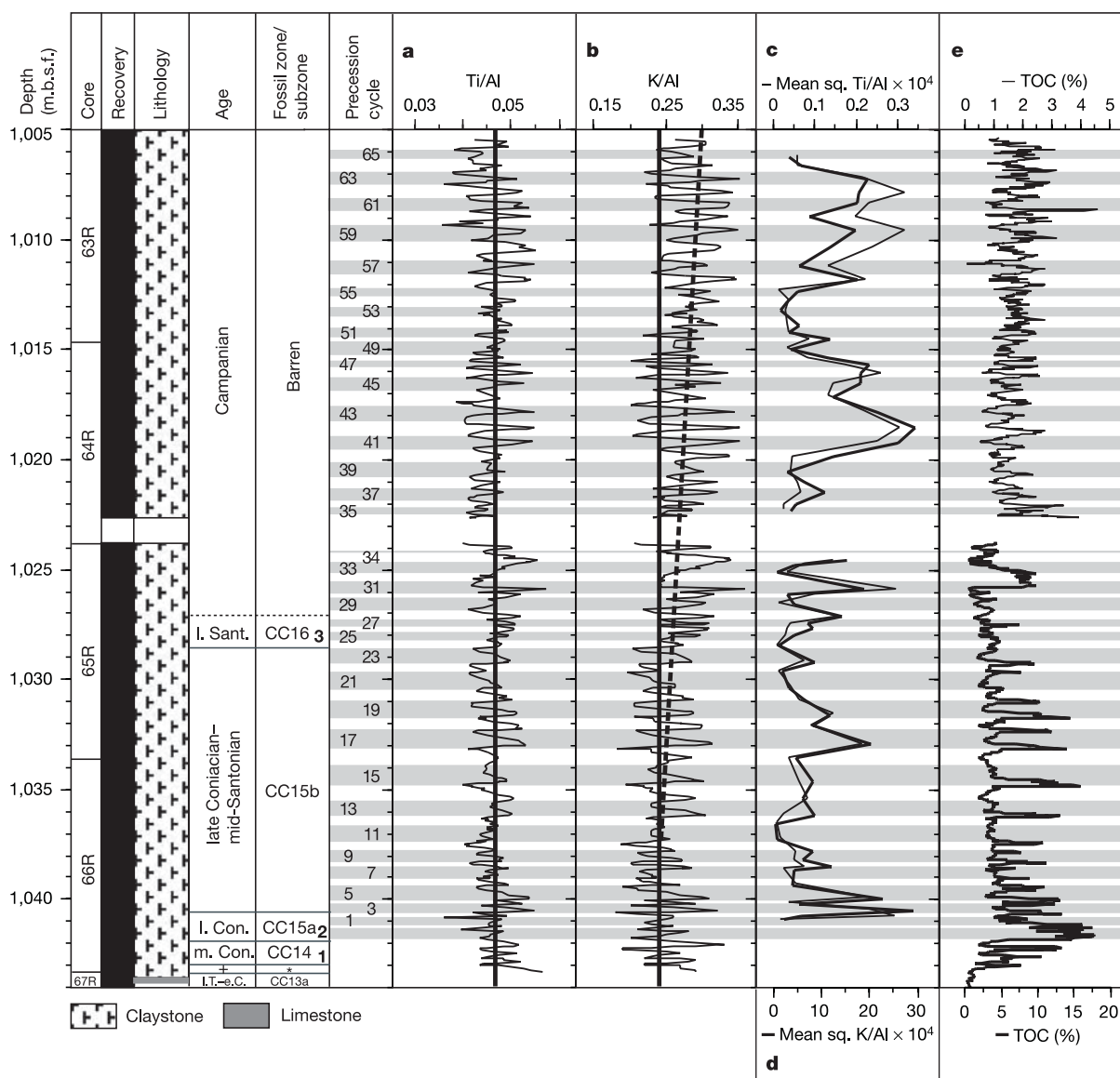


Figure 2 | Upper Cretaceous (Coniacian to Campanian) records from ODP Site 959. **a**, Ti/Al ratio, average value indicated by solid black line. **b**, K/Al ratio, average value for lower part of the profile indicated by solid line, mean values starting from 1,034 m.b.s.f., marked by dashed line. Both proxies (**a**, **b**) serve as a measure of fluctuations between tropical and arid climate conditions; higher values indicate greater aridity. **c**, Mean square Ti/Al values (thin line). **d**, Mean square K/Al values (heavy line). Both proxies (**c**, **d**) are calculated as the standard deviation from respective mean values for each precession cycle; higher values indicate greater seasonality, that is,

more marked alternations between arid and humid conditions. **e**, Total organic carbon content (TOC) in %; bottom axis refers to values below 1,022.5 m.b.s.f. (heavy line); top axis refers to data above 1,022.5 m.b.s.f. (thin line). Higher values (15–20%) indicate anoxia. Fossil zone/subzone and age from refs 21, 22; I. Sant., late Santonian; I. Con., late Coniacian; m. Con., mid Coniacian; I. T.–e. C., late Turonian to early Coniacian; *CC13b; + early Coniacian; absolute ages²³; 1, 87.2 Myr; 2, 85.5 Myr; 3, 84.8 Myr. Precession cycles (numbered from bottom to top) below 1,024 m.b.s.f. are based on previous work^{6,8}.

time control and sedimentation rate at Site 959, we conclude that this prominent periodicity lies close enough to 22 kyr to support the influence of precession. (Details of the time–frequency analysis are available as Supplementary Information.) Given the persistence of the dominant periodicity with respect to total organic carbon (TOC) burial across the core break above 1,022.5 metres below the sea floor (m.b.s.f.), we estimate 65 cycles for the entire section (1,044–1,005 m.b.s.f.), equivalent to ~ 1.46 Myr. Variance records of K/Al and Ti/Al reveal the evolution of African climate contrasts (Fig. 2c, d). High values document alternations between arid and humid conditions, and low values indicate more balanced climate. Above 1,022.5 m.b.s.f. a robust pattern consisting of 19 cycles emerges (Fig. 2c, d), corresponding to the ~ 400 kyr period. The variance records below 1,022.5 m.b.s.f. show a frequency with a periodicity of ~ 315 kyr (~ 14 cycles in total). This regularity is not attributable to the main orbital frequencies, and cannot easily be explained. We consider the manifestation of a 400 kyr periodicity to be the expression of a long-term drying/cooling cycle. The presence of long eccentricity and precession cycles supports Upper Cretaceous climate as having been dynamic and variable, comparable to that of the Quaternary.

A linkage of African continental climate with OC burial and oceanic productivity on a precession cycle level has been reported before^{6–8}, and is evident when the Coniacian–Campanian OC record from Site 959 is compared to the Ti/Al and K/Al profiles. The section below 1,024 m.b.s.f. displays a distinct cyclic, high-amplitude pattern⁸ that fades out towards the top owing to the progressive subsidence of Site 959 through time^{14,15} (Fig. 2e). Despite the decline in amplitudes, maxima in OC generally coincide with minima in K/Al and Ti/Al, that is, with periods of enhanced detrital input from tropical source areas.

OC accumulation within individual cycles exhibits a distinct pattern (Fig. 2e) characterized by an initial sharp increase followed by a plateau and a gradual decrease. Only peak OC deposition was associated with the development sea floor anoxia and photic zone euxinia⁸. A compilation of OC data from Late Coniacian–Mid Santonian nannofossil zone CC15 reveals that maximum black shale sedimentation on average lasted for a quarter cycle, that is, ~ 5 kyr (ref. 7). The robustness of this pattern across the record points to an external control that caused episodic anoxic conditions in the tropical eastern Atlantic over more than one million years. A previous investigation⁶ proposed that black shale formation was triggered by discharge of freshwater and nutrients to the

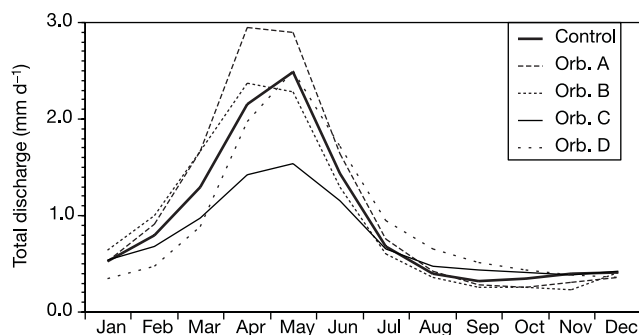


Figure 3 | Simulated total continental freshwater discharge over the course of one precession cycle (orbital cases A to D) compared to control run conditions. Total continental freshwater discharge (TD) was drastically enhanced during spring (April–May), independent of orbital configuration. Highest levels of up to 3 mm d^{-1} were achieved during orbital configuration A (warm spring/cold autumn in the Northern Hemisphere), in parallel with strongest seasonal contrasts. From orbital case A to C climate contrast diminishes, resulting in more balanced cold spring/warm autumn conditions and a decrease in TD by about a factor of 2 during orbital case C. Orbital cases B and D represent transitional phases.

semi-enclosed DIB, resulting in a circulation reversal and the development of ocean anoxia/euxinia^{6,8}. Although less well constrained, a similar mechanism and long-term climate trend has been suggested for the South American Maracaibo basin during OAE 3 (ref. 16).

The implications from the records of Site 959 were validated using global climate model GENESIS v.2.0 (ref. 17). (Details of the model and the experimental set-up of GENESIS v. 2.0 are available as Supplementary Information.) Five simulations with changing orbital configurations were conducted: a control run with no precession and four runs covering one precessional cycle, each shifting the closest annual approach of the Earth to the Sun (perihelion) into a different season.

Of specific relevance to the black shale mode is total freshwater discharge (TD). Previous simulations emphasized the importance of continental discharge for climate change and ocean response; for example, during the ice-free Early Eocene¹⁸. All simulations performed in this study yielded a pronounced wet season in spring and early summer (March to June, Fig. 3) with enhanced TD accounting for 56–71% of the total annual discharge (Table 1), and a dry season (July to February). Strongest seasonal contrasts occurred during orbital case A, representing the orbital configuration when spring is coincident with perihelion. Weakest contrasts are recognized for case C, suggesting most balanced conditions when perihelion coincided with autumn configuration. The two remaining orbital cases (B and D) describe ‘transitional’ modes. During case A, total annual TD was elevated by about 43% relative to case C. During the months of wettest conditions and largest divergence (March–June) TD was elevated by 79% ($\sim 487 \text{ km}^3 \text{ yr}^{-1}$), but reduced by 27% ($\sim 105 \text{ km}^3 \text{ yr}^{-1}$) during dry seasons (July–February) for case A relative to case C (Table 1).

The fluctuations in TD for orbital configurations A to D provide a robust mechanism to explain the observed pattern in ocean redox conditions and OC burial. Black shale formation apparently was restricted to case A, when maximum insolation (northern spring equinox at perihelion) occurred during wet seasons, both fostering seasonal contrasts and massive spring TD (Fig. 3). These conditions apparently were not reached during orbital configurations B, C and D. The runoff data suggest that a threshold in excess of $125 \text{ km}^3 \text{ yr}^{-1}$ was necessary to maintain a year-round or multi-year freshwater cap that triggered a reversal in ocean circulation in the DIB and a shift into a black shale mode. The estimated thickness of a freshwater cap ranges from 1.0 m during orbital case A to 0.7 m during orbital case C (Table 1), providing values comparable to the modern Black Sea (0.7 m; <http://www.grid.unep.ch/bsein/index.html>) but 2–3 times higher than for the modern Arctic Ocean (0.36 m)¹⁹. The direct linkage of African TD patterns and black shale formation in the DIB implies that fluctuations in ocean properties were short, that is, on seasonal timescales, but had a major effect on the Upper Cretaceous carbon cycle.

The results of this study demonstrate how sensitively and rapidly tropical marine areas close to continental margins react to even relatively moderate increases in continental freshwater discharge. The freshwater threshold required to shift sheltered and semi-enclosed areas of the modern ocean into an anoxic mode are unknown, but the progressive emission of greenhouse gases to the modern atmosphere is gradually shifting Earth towards a greenhouse mode with an accelerated hydrological cycle. At present it is hardly possible to estimate where we are on that long-term climate trend, but once the freshwater threshold is passed, a substantial impact on biochemical cycling of continental margins may be expected.

METHODS

Original data. Site 959 was drilled during ODP Leg159 off the Ivory Coast and Ghana at $3^\circ 37.656' \text{ N}$, $2^\circ 44.149' \text{ W}$ in approximately 2,100 m water-depth. Samples were taken continuously in 1 cm increments, dried at 35°C and homogenized before analysis. Determination of OC was performed after

removal of inorganic carbon with 0.25 N HCl on every other sample. Major element chemistry was determined by X-ray fluorescence analysis on samples with an average spacing of every 8 cm.

Cycle and age model. The cyclostratigraphic model for the record of Hole 959D between 1,005.43 and 1,046.09 m.b.s.f. is based on the analysis of geochemical cycles within the interval below 1,022.5 m.b.s.f., in particular biozone CC15b⁶. In this biozone, the number of cycles in the Si/Al and K/Al records were matched to precession by counting the number of cycles in a biostratigraphically well-defined time interval. The cycle pattern with respect to the Si/Al, Ti/Al, K/Al and TOC records is not restricted to biozone CC15b, but clearly extends over the entire interval between 1,005.43 and 1,046.09 m.b.s.f. The biostratigraphic time control above 1,022.5 m.b.s.f., however, prohibits extending the approach described for biozone CC15b. Time–frequency analysis in CC15 reveals one prominent frequency of more than 3 cycles m⁻¹ at the base that shifts to about 1.3 cycles m⁻¹ above 1,037 m.b.s.f. The upper part of the profile between 1,022.69 and 1,005.43 m.b.s.f. also shows one prominent frequency at about 1.8 cycles m⁻¹ that progressively decreases upward to about 1.3 cycles m⁻¹. No persistent signal components are detected at shorter wavelengths. The period of the dominant cycle in the time domain was estimated by analysing the data from nannofossil biozone CC15 above 1,041 m.b.s.f. in a chronostratigraphic framework. Time–frequency analysis for TOC and K/Al displays one prominent frequency at about 35 cycles Myr⁻¹ throughout the main parts of the record. Based on the age model, this frequency corresponds to a period of about 28 kyr; this is close enough to 22 kyr to support precession as the prominent frequency. See Supplementary Information for further details.

Global circulation model. The general features of the GENESIS (Global Environmental Ecological Simulation of Interactive Systems) 2.0, designed for palaeoclimate research, have been presented before¹³. Boundary conditions were set for Cenomanian–Turonian simulations (see Supplementary Information for specifications of experimental set-up, palaeogeography, astronomical forcing, and validation of GENESIS v. 2.0). Five simulations with changing orbital forcing were carried out.

Received 2 June; accepted 28 June 2005.

- Wilson, P. A. & Norris, R. D. Warm tropical ocean surface and global anoxia during the mid-Cretaceous period. *Nature* **412**, 425–429 (2001).
- Norris, R. D., Bice, K. L., Magno, E. A. & Wilson, P. A. Jiggling the tropical thermostat in the Cretaceous hothouse. *Geology* **30**, 299–302 (2002).
- Dean, W. E., Arthur, M. A. & Stow, D. A. V. Origin and geochemistry of Cretaceous black shales and multicoloured claystones, with emphasis on Deep Sea Drilling Project Site 530, southern Angola Basin. *DSDP Init. Rep.* **75**, 819–844 (1984).
- Erlich, R. N., Palmer-Koleman, S. E. & Antonieta Lorente, M. Geochemical characterization of oceanographic and climatic changes recorded in upper Albian to lower Maastrichtian strata, western Venezuela. *Cretac. Res.* **20**, 547–581 (1999).
- Rusk, D. C. in *Petroleum Provinces of the Twenty-first Century* (eds Downey, M. W., Threet, J. C. & Morgan, W. A.) 429–452 (Memoir 74, AAPG, Tulsa, 2001).
- Hofmann, P., Wagner, T. & Beckmann, B. Millennial- to centennial-scale record of African climate variability and organic carbon accumulation in the Coniacian-Santonian eastern tropical Atlantic (Ocean Drilling Program Site 959, off Ivory Coast and Ghana). *Geology* **31**, 135–138 (2003).
- Beckmann, B., Wagner, T. & Hofmann, P. in *The Deposition of Organic-Carbon-Rich Sediments: Models, Mechanisms, and Consequences* (ed. Harris, N. B.) 125–143 (Spec. Publ. 82, SEPM, Washington, 2005).
- Wagner, T., Sinninghe Damsté, J. S., Hofmann, P. & Beckmann, B. Euxinia and primary production in Late Cretaceous eastern equatorial Atlantic surface waters fostered orbitally driven formation of marine black shales. *Paleoceanography* **19**, doi:10.1029/2003PA000898 (2004).
- Calvert, S. E. in *Treatise on Chemical Oceanography* 6 (eds Riley, J. P. & Chester, R.) 187–280 (Academic, San Diego, 1976).
- Zabel, M. et al. Late Quaternary climate changes in Central Africa as inferred from terrigenous input to the Niger Fan. *Quat. Res.* **56**, 207–217 (2001).
- Diester-Haass, L. & Chamley, R. C. Palaeoceanographic and palaeoclimatic evolution in the Wedell Sea (Antarctica) during the middle Eocene-late Oligocene, from a coarse sediment fraction and clay mineral data (OCP 689). *Mar. Geol.* **114**, 233–250 (1993).
- Yarincik, K. M., Murray, R. W. & Peterson, L. C. Climatically sensitive eolian and hemipelagic deposition in the Cariaco Basin, Venezuela, over the past 578,000 years: Results from Al/Ti and K/Al. *Paleoceanography* **15**, 210–228 (2000).
- DeConto, R. M., Hay, W. W., Thompson, S. L. & Bergengren, J. in *The Evolution of the Cretaceous Ocean-Climate System* (eds Barrera, E. & Johnson, C.) 391–406 (Spec. Pap. 332, Geol. Soc. Am., Boulder, 1999).
- Holbourn, A. E. L. et al. in *The Oil and Gas Habitats of the South Atlantic* (eds Cameron, N. R., Bate, R. H. & Clure, V. S.) 195–222 (Spec. Publ. 153, Geol. Soc., London, 1999).
- Wagner, T. Late Cretaceous to early Quaternary organic sedimentation in the eastern equatorial Atlantic. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **179**, 113–147 (2002).
- Rey, O., Simo, J. A. & Lorente, M. A. A record of long- and short-term environmental and climatic change during OAE3: La Luna Formation, Late Cretaceous (Santonian-early Campanian), Venezuela. *Sedim. Geol.* **170**, 85–105 (2004).
- Thompson, S. L. & Pollard, D. A global climate model (GENESIS) with a land-surface-transfer scheme (LSX). Part 1: Present-day climate. *J. Clim.* **8**, 732–761 (1995).
- Bice, K. L., Barron, E. J. & Peterson, W. H. Continental runoff and early Cenozoic bottom-water sources. *Geology* **25**, 951–954 (1997).
- Alychinskaya, N. M. & Ivanov, V. V. in *World Water Balance and Water Resources of the Earth* (ed. Korzun, V. I.) 568–574 (Studies and Reports in Hydrology 25, UNESCO, Paris, 1978).
- Hay, W. W., et al. in *The Evolution of the Cretaceous Ocean-climate System* (eds Barrera, E. & Johnson, C.) 1–48 (Spec. Pap. 332, Geol. Soc. Am., Boulder, 1999).
- Watkins, D. K., Shafik, S. & Shin, I. C. Calcareous nannofossils from the Cretaceous of the Deep Ivory Basin. *ODP Sci. Res.* **159**, 319–333 (1998).
- Kuhnt, W., Moullade, M. & Kaminski, M. A. Upper Cretaceous, K/T boundary, and Paleocene agglutinated foraminifers from Hole 959D (Côte d'Ivoire-Ghana Transform Margin). *ODP Sci. Res.* **159**, 389–411 (2004).
- Shipboard Scientific Party. Explanatory notes. *Proc. ODP Init. Rep.* **207**, 1–94 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the Ocean Drilling Program for providing the sample material and technical assistance. H. Heilmann, R. Henning, B. Kockisch, M. Gölden, S. Berg, A. Katzemich, S. Reiss, M. Aiyaswaryan, R. Kunze and D. Pollard are thanked for analytical assistance.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.W. (thomas.wagner@ncl.ac.uk).

Carbon losses from all soils across England and Wales 1978–2003

Pat H. Bellamy¹, Peter J. Loveland¹, R. Ian Bradley¹, R. Murray Lark² & Guy J. D. Kirk¹

More than twice as much carbon is held in soils as in vegetation or the atmosphere¹, and changes in soil carbon content can have a large effect on the global carbon budget. The possibility that climate change is being reinforced by increased carbon dioxide emissions from soils owing to rising temperature is the subject of a continuing debate^{2–9}. But evidence for the suggested feedback mechanism has to date come solely from small-scale laboratory and field experiments and modelling studies^{2–9}. Here we use data from the National Soil Inventory of England and Wales obtained between 1978 and 2003 to show that carbon was lost from soils across England and Wales over the survey period at a mean rate of $0.6\text{ g kg}^{-1}\text{ yr}^{-1}$ (relative to the existing soil carbon content). We find that the relative rate of carbon loss increased with soil carbon content and was more than $2\% \text{ yr}^{-1}$ in soils with carbon contents greater than 100 g kg^{-1} . The relationship between rate of carbon loss and carbon content is irrespective of land use, suggesting a link to climate change. Our findings indicate that losses of soil carbon in England and Wales—and by inference in other temperate regions—are likely to have been offsetting absorption of carbon by terrestrial sinks.

The National Soil Inventory was made to obtain an unbiased estimate of the distribution of the soils of England and Wales and of the chemistry of the topsoil (0–15 cm depth)¹⁰. Samples were collected and soil profiles described at the intersections of an orthogonal 5-km grid over the whole area (Methods). This yielded about 6,000 sites, of which 5,662 could be sampled for soil. Figure 1a

shows the distribution of soil organic carbon contents across England and Wales measured in the original sampling (1978–83). Sufficient subsets of the sites were resampled at intervals from 12 to 25 yr after the original sampling to be able to detect changes in carbon content with 95% confidence (Methods). This was done in three phases: in 1994–95 for arable and rotational grassland sites (853 of the original 2,578 sites), in 1995–96 for managed permanent grassland sites (771 of the original 1,579), and in 2003 for non-agricultural sites (bogs, scrub, rough grazing, woodland, and so on; 555 of the original 1,505). Roughly 40% of the original sites were resampled. This is the only soil inventory on such a scale anywhere in the world to have been resampled. To allow for the varying time interval between samplings, annual rates of change in carbon were calculated for each site by assuming that the process of change was linear over the sampling interval. An analysis of known rates of change in soil carbon under different conditions showed this to be reasonable.

Figure 2 summarizes the results grouped by soil type and land use. Some differences between soils and land uses are apparent: for example, peat soils lost carbon an order of magnitude faster than brown soils and man-made soils, and bogs and upland grass lost carbon an order of magnitude faster than lowland heath, which appears to have gained carbon on average. But we found no statistically significant relations between rate of change and land use, rainfall class or soil textural class, whether for the data as a whole or for outlying data. However, we found a significant negative linear correlation between rate of change and original organic carbon

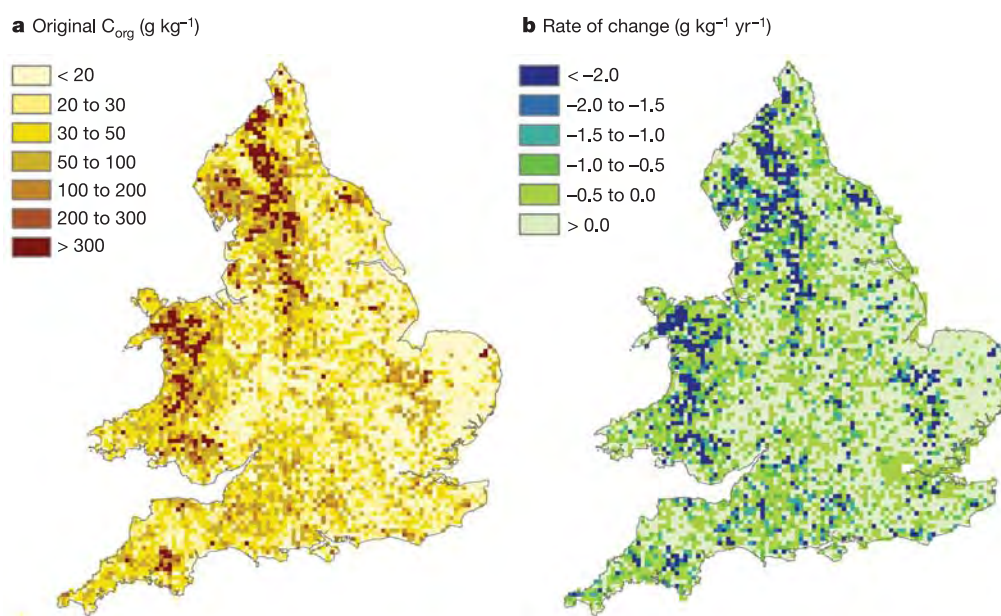


Figure 1 | Changes in soil organic carbon contents across England and Wales between 1978 and 2003. **a**, Carbon contents in the original samplings, and **b**, rates of change calculated from the changes over the different sampling intervals. Values at sites that were not resampled were calculated from their original organic carbon contents using equation (1). The changes were negative in all but 8% of the sites.

¹National Soil Resources Institute, Cranfield University, Silsoe MK45 4DT, UK. ²Rothamsted Research, Harpenden AL5 2JQ, UK.

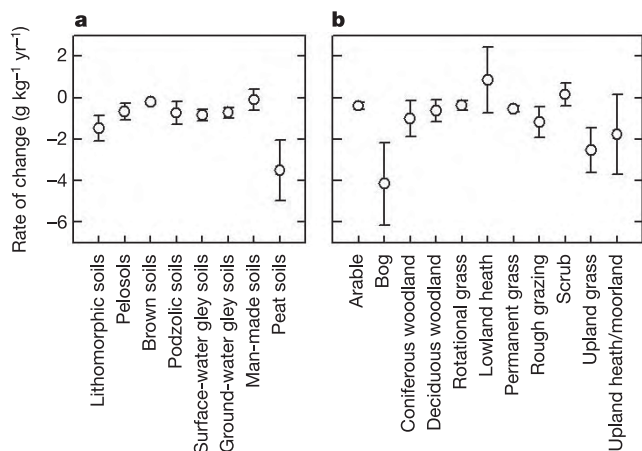


Figure 2 | Rates of change in soil organic carbon content, grouped by soil type and land use. a, Soil type²⁴ grouping; b, land use grouping. Circles indicate mean values; error bars indicate 95% confidence intervals.

content (C_{org}); that is, the rate of loss increased with C_{org} (Fig. 3). The relation applied for the data as a whole and for the three main land use groupings separately, though with different slopes and intercepts. We took a random subsample of 1,000 observations from the data as a whole, and obtained the following relation by linear regression specifying an exponential variogram for the residuals (Methods):

$$\text{Rate of change in } C_{org} = 0.6 - 0.0187 \times \text{original } C_{org} \quad (1)$$

The standard error (s.e.) of the intercept is 0.148 and the s.e. of the slope is 0.00081; the rate of change in C_{org} is in units of $\text{g kg}^{-1} \text{yr}^{-1}$, and C_{org} is in g kg^{-1} . The residuals from this regression applied to the whole data set show some regional features, notably a tendency to overestimate the rate of loss in uplands in the northwest and southwest of England. But only 8% of the unexplained variation is spatially structured, so a more sophisticated statistical model is unjustified. Figure 1b shows the distribution of rates of change across England and Wales for all the sites, with values for the sites that were not resampled obtained using equation (1).

From the data in Fig. 3, the relative rates of change (rate of change/mean C_{org} over sampling interval, in units of $\% \text{yr}^{-1}$) were 1.43, 0.14, -0.69, -1.84, -2.71, -3.01 and -2.35 for original C_{org} ranges (in g kg^{-1}) 0–20, 20–30, 30–50, 50–100, 100–200, 200–300 and >300, respectively. Hence the relative rates of loss also tended to increase with C_{org} , although not above 300 g kg^{-1} . Soils with $C_{org} < 50 \text{ g kg}^{-1}$ did not lose significant amounts of carbon (that is, not detectable over 10 yr), and those with $C_{org} < 20 \text{ g kg}^{-1}$ appear to have gained it. However, for soils with $C_{org} > 100 \text{ g kg}^{-1}$, relative rates of loss over the survey period were $>2\% \text{yr}^{-1}$. Given that the bulk of the

UK's carbon stocks are in organic soils¹¹, this result gives cause for concern.

Table 1 gives estimates of the total changes in carbon in the upper 15 cm of soil for the whole of England and Wales, with rates for the sites that were not resampled predicted using equation (1). The total amount of carbon in the upper 15 cm at the time of the original sampling is estimated to be 864 Tg, and the total rate of change in this depth is -4.44 Tg yr^{-1} . From the distribution of soil types in the 1:250,000 National Soil Map interpolated on a 1-km grid, Bradley *et al.*¹¹ estimate the total carbon stock of the soils of the UK to be 2,542 Tg over 0–30 cm depth (1,015 Tg in England, 194 Tg in Wales, 1,161 Tg in Scotland, 172 Tg in Northern Ireland). If the rate of change found here also applies to the soils of Scotland (but note that most soils in Scotland have large organic carbon contents¹¹, and therefore the true rate of loss is likely to be larger), and if the rate of change over 0–30 cm depth is the same as over 0–15 cm, then *pro rata* the total rate of loss across the UK is 13 Tg yr^{-1} . For comparison, the UK's current industrial CO_2 emission is about 150 Tg yr^{-1} .

We cannot with the existing data say where the missing carbon has gone. Some proportion will have been emitted to the atmosphere as CO_2 and some will have been leached to deeper soil layers and to drainage waters and beyond. The latter would be consistent with the observed increases in dissolved organic carbon in surface waters across much of England and Wales over the past 40 yr (refs 12, 13).

Soil carbon contents depend on rates of addition from plant growth versus rates of removal in decomposition, leaching and other soil processes, and each of these is sensitive to changes in land use, climate and other variables^{14–17}. Various changes in land use will have contributed to carbon losses from soils across England and Wales over the survey period, both under agricultural uses (drainage schemes, post-war grassland conversion, increased stocking rates) and non-agricultural uses (afforestation on wet soils, increased erosion, increased burning of upland vegetation). However, we do not have sufficient data at the scale of the National Soil Inventory to explore these effects.

The fact that the losses appear to be happening across both countries irrespective of land use suggests a link to climate change. Over the survey period, the mean temperature across England and Wales increased by about 0.5°C and there were also changes in rainfall distribution¹⁸. Climate change will affect soil carbon turnover through various processes. Increases in temperature will tend to increase rates of organic matter decomposition by soil microbes, although the magnitude of this effect and differences between soils are uncertain⁹. The effects of temperature will interact in a complicated way with changes in soil moisture brought about by changing rainfall and evapo-transpiration patterns, and changes in atmospheric CO_2 and nitrogen deposition. In freely drained soils, warmer drier conditions may retard decomposition of organic matter if lack of moisture limits soil microbes. But in wet anoxic soils, increased

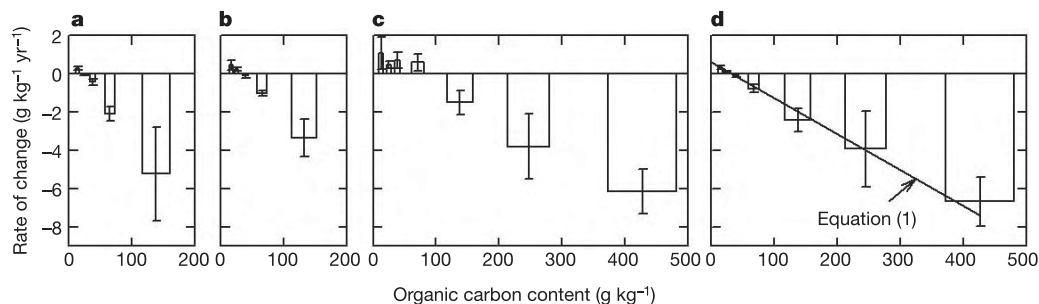


Figure 3 | Rates of change in soil organic carbon content, grouped by original carbon contents and indicated land uses. a, Arable/rotational grass; b, permanent grass; c, non-agricultural; d, all. The ranges of carbon content (g kg^{-1}) are, from left to right in each panel, 0–20, 20–30, 30–50, 50–100, 100–200, 200–300 and >300. Bar heights indicate mean values; error bars indicate 95% confidence intervals. The derivation of equation (1) is described in the text.

100, 100–200, 200–300 and >300. Bar heights indicate mean values; error bars indicate 95% confidence intervals. The derivation of equation (1) is described in the text.

Table 1 | Estimated rates of change for all original sites

Range of original C_{org} ($g\ kg^{-1}$)	No. sites originally sampled	Mean original C_{org} ($g\ kg^{-1}$)	Total area		Total C_{org} 0–15 cm depth*		Rate of change in C_{org} † ($g\ kg^{-1}\ yr^{-1}$)	Total change in 0–15 cm‡ ($Tg\ yr^{-1}$)
			(km^2)	(%)	(Tg)	(%)		
0–20	1,061	13.9	26,525	18.7	66.4	7.7	0.34	1.68
20–30	1,158	24.5	28,950	20.5	111.9	12.9	0.13	0.66
30–50	1,607	38.5	40,175	28.4	214.8	24.9	–0.10	–0.65
50–100	1,140	66.8	28,500	20.1	220.5	25.5	–0.68	–2.11
100–200	313	137.6	7,825	5.5	92.5	10.7	–2.18	–1.32
200–300	95	244.5	2,375	1.7	36.5	4.2	–4.00	–0.59
>300	288	439.7	7,200	5.1	121.7	14.1	–7.37	–2.10
All	5,662	66.1	141,550	100	864.3	100	–0.64	–4.44

* Original C_{org} $\times$ bulk density ($kg\ dm^{-3}$) $\times$ depth (dm) $\times$ area (km^2) $\times 10^{-4}$, where bulk density = $1.3 - (0.275 \ln(C_{org}/10))$ (ref. 25).

† For sites not resampled, rate of change was calculated from original C_{org} using equation (1).

‡ Rate of change $\times$ bulk density $\times$ area.

oxygenation of the soil as evaporation increases the depth to the water table will greatly reinforce the increase in decomposition with increased temperature. Such effects may in part explain the increase in the relative rate of loss with soil carbon content because more-organic soils tend to be wetter. Also consistent with this, more-organic soils necessarily contain a greater proportion of slowly decaying organic matter, and the rate of turnover of this material appears to be more sensitive to temperature changes than that of more-labile organic matter⁸. However the magnitudes of these effects are unknown.

On the basis of atmospheric observations, net carbon absorption by terrestrial systems in the Northern Hemisphere has increased in recent decades¹⁴. The magnitude of the carbon sink in different regions and the contributions of different processes are highly uncertain, but the main sinks are thought to be changes in land use and increased forest growth with increased atmospheric CO_2 and nitrogen deposition. Our findings show that losses of soil carbon in the UK, and by inference in other temperate regions, are likely to have been offsetting absorption by terrestrial sinks, greatly adding to the uncertainty of future trends.

METHODS

Original sampling. The 5-km sampling grid was offset 1 km north and 1 km east of the origin of the National Grid to avoid sampling at the edges of published maps. Urban areas and water bodies were avoided, but otherwise all soils were sampled. At each site, 25 soil cores were taken at 4-m intervals in a 20 m $\times$ 20 m square centred on the grid intersection¹⁹. The cores were taken to a maximum depth of 15 cm from the surface using a 2.5-cm diameter auger, with the soil surface taken as the zero of measurement, and excluding litter layers and living vegetation. The cores were bulked in the field, giving a total of approximately 1 kg of moist material for each site. The samples were air-dried at temperatures not exceeding 30°C, and each then divided into three equal portions. One portion was stored in a polythene bag inside a cardboard box at ambient temperature and humidity without further treatment. The other portions were crushed to <2 mm, and subsamples analysed for organic carbon as below.

Resampling. The minimum proportions of the original sites to be resampled so as to detect changes in C_{org} with 95% confidence were calculated using the relation $n_2/n_1 \geq 1/[1 + n_1(d/z_\alpha)^2]$, where n_1 and n_2 are the sizes of the original and resampled populations, α is the probability that the change in mean C_{org} is greater than some small value d , z_α is the probability of the standardized normal distribution of C_{org} being less than α , and s is the standard deviation of the original population. The accuracy of the C_{org} measurement in the laboratory (see below) was $\pm 1\ g\ kg^{-1}$, and hence the minimum value of d is $2\ g\ kg^{-1}$. This value was used for the arable/rotational grass and permanent grass sites. For the non-agricultural sites, the standard deviation of C_{org} was greater, particularly for more-organic soils. Accordingly, for these sites sampling was designed with $d = 2\ g\ kg^{-1}$ for soils with $C_{org} < 150\ g\ kg^{-1}$, $d = 5\ g\ kg^{-1}$ for $C_{org} = 150–300\ g\ kg^{-1}$ and $d = 10\ g\ kg^{-1}$ for $C_{org} > 300\ g\ kg^{-1}$. Sites were selected at random from the original sites within these groupings in proportion to the original regional sampling density, and sampled exactly as originally.

To test the accuracy with which the sites could be relocated, six surveyors were instructed to revisit 10 sites with widely differing characteristics following the originally recorded site descriptions, their positions recorded with a global positioning system, and the distances between them subsequently measured.

This showed the accuracy of relocation was better than 20 m on enclosed land and better than 50 m on open land, and values of C_{org} at 0, 20 and 50 m from the target measured using the original methods were not significantly different from each other (Supplementary Fig. 1). Hand-written field records from the original sampling were examined to see if outliers in the data (based on the log normalized changes in C_{org}) had any common features. No artefacts were found that would have led us to believe the outliers were not true representations of the changes.

Organic carbon analysis. Exactly the same methods were used for the two samplings. Soils with $C_{org} < 150\ g\ kg^{-1}$ were analysed by a modified Walkley-Black method²⁰. The small subsamples this uses are unrepresentative of highly organic soils, so soils with $C_{org} > 150\ g\ kg^{-1}$ were analysed by loss on ignition (LOI)²¹, converted to C_{org} by $C_{org} = 0.5 \times LOI$. To check that the change of method introduced no artefacts, we applied both to 95 soils with $C_{org} = 20–200\ g\ kg^{-1}$ and obtained good agreement and no systematic deviation (Supplementary Fig. 2). To check for differences in analytical precision between the samplings, we reanalysed stored samples from 10% of the original sites and obtained good agreement with the original values across the full range of C_{org} (Supplementary Fig. 3).

Derivation of equation (1). Equation (1) is a regression, but was not fitted by ordinary least squares because the original systematic sampling precludes the assumption that the residuals are independent random variables²². Equation (1) was fitted as a mixed model with a spatially-dependent random effect and a white noise term such that, for any pair of locations x_i and x_j , the expected squared difference of the residuals, $\eta(x_i)$ and $\eta(x_j)$, is $E[\{\eta(x_i) - \eta(x_j)\}^2] = 2\{c_0 + c_1(1 - \exp[-|x_i - x_j|/a])\}$, where c_0 and c_1 are the variances of the white noise and spatially dependent components, respectively, and a is a distance parameter. The model was fitted by residual maximum likelihood²³.

Received 24 February; accepted 8 July 2005.

- Batjes, N. H. Total carbon and nitrogen in the soils of the world. *Eur. J. Soil Sci.* **47**, 151–163 (1996).
- Jenkinson, D. S., Adams, D. E. & Wild, A. Model estimates of CO_2 emissions from soil in response to global warming. *Nature* **351**, 304–306 (1991).
- Cao, M. K. & Woodward, F. I. Dynamic responses of terrestrial ecosystem carbon cycling to global climate change. *Nature* **393**, 249–252 (1998).
- Giardina, C. P. & Ryan, M. G. Evidence that decomposition rates of organic carbon in mineral soil do not vary with temperature. *Nature* **404**, 858–861 (2000).
- Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**, 184–187 (2000).
- Davidson, E. A., Trumbore, S. E. & Amundson, R. Soil warming and organic carbon content. *Nature* **408**, 789–790 (2000).
- Fang, C., Smith, P., Moncrieff, J. B. & Smith, J. U. Similar response of labile and resistant soil organic matter pools to changes in temperature. *Nature* **433**, 57–59 (2005).
- Knorr, W., Prentice, I. C., House, J. I. & Holland, E. A. Long-term sensitivity of soil carbon turnover to global warming. *Nature* **433**, 298–301 (2005).
- Powlson, D. S. Will soil amplify climate change? *Nature* **433**, 204–205 (2005).
- Loveland, P. J. in *Element Concentration Cadasters in Ecosystems* (eds Lieth, H. & Markert, B.) 73–80 (VCH, Weinheim, Germany, 1990).
- Bradley, R. I. et al. A soil carbon and land use database for the UK. *Soil Use Manag.* (in the press).
- Worrall, F., Burt, T. & Shedden, R. Long term records of riverine dissolved organic matter. *Biogeochemistry* **64**, 165–178 (2003).
- Freeman, C. et al. Export of dissolved organic carbon from peatlands under elevated carbon dioxide levels. *Nature* **430**, 195–198 (2004).

14. Schimel, D. S. *et al.* Recent patterns and mechanisms of carbon exchange by terrestrial ecosystems. *Nature* **414**, 169–172 (2001).
15. Guo, L. B. & Gifford, R. M. Soil carbon stocks and land use change: a meta-analysis. *Glob. Change Biol.* **8**, 345–360 (2002).
16. Houghton, R. A. Revised estimates of the annual net flux of carbon to the atmosphere from changes in land use and land management 1850–2000. *Tellus B* **55**, 378–390 (2003).
17. Smith, K. A. *et al.* Exchange of greenhouse gases between soil and atmosphere: interactions of soil physical factors and biological processes. *Eur. J. Soil Sci.* **54**, 779–791 (2003).
18. Hulme, M. *et al.* *Climate Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report* (Tyndall Centre for Climate Change Research, Norwich, UK, 2002).
19. Webster, R. & Burgess, T. M. Sampling and bulking strategies for estimating soil properties in small regions. *J. Soil Sci.* **35**, 127–140 (1984).
20. Kalembasa, S. J. & Jenkinson, D. S. A comparative study of titrimetric and gravimetric methods for the determination of organic carbon in soil. *J. Sci. Food Agric.* **24**, 1085–1090 (1973).
21. Avery, B. W. & Bascomb, C. L. *Soil Survey Laboratory Methods* (Soil Survey Technical Monograph No. 6, Soil Survey of England & Wales, Harpenden, UK, 1974).
22. Lark, R. M. & Cullis, B. R. Model-based analysis using REML for inference from systematically sampled data on soil. *Eur. J. Soil Sci.* **55**, 777–797 (2004).
23. Gilmour, A. R. *et al.* ASReml User Guide release 1.0. (VSN International, Hemel Hempstead, UK, 2002).
24. Avery, B. W. *Soil Classification for England and Wales (Higher Categories)* (Soil Survey Technical Monograph No. 14, Soil Survey of England & Wales, Harpenden, 1980).
25. Howard, P. J. A. *et al.* The carbon content of soil and its geographical distribution in Great Britain. *Soil Use Manag.* **11**, 9–15 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank Defra for funding this research, R. Andrews for technical assistance, and D. Powlson and J. Hollis for comments on the draft Letter.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to G.J.D.K. (g.kirk@cranfield.ac.uk).

Geophysical evidence from the MELT area for compositional controls on oceanic plates

Rob. L. Evans¹, Greg Hirth¹, Kiyoshi Baba³, Don Forsyth⁴, Alan Chave² & Randall Mackie⁵

Magnetotelluric and seismic data, collected during the MELT experiment at the southern East Pacific Rise^{1,2}, constrain the distribution of melt beneath this mid-ocean-ridge spreading centre and also the evolution of the oceanic lithosphere during its early cooling history. Here we focus on structures imaged at distances ~ 100 to 350 km east of the ridge crest, corresponding to seafloor ages of ~ 1.3 to 4.5 million years (Myr), where the seismic and electrical conductivity structure is nearly constant and independent of age. Beginning at a depth of about 60 km, we image a large increase in electrical conductivity and a change from isotropic to transversely anisotropic electrical structure, with higher conductivity in the direction of fast propagation for seismic waves. Conductive cooling models predict structure that increases in depth with age, extending to about 30 km at 4.5 Myr ago. We infer, however, that the structure of young oceanic plates is instead controlled by a decrease in water content above a depth of 60 km induced by the melting process beneath the spreading centre³.

MELT magnetotelluric response functions were inverted for a transversely anisotropic conductivity structure, described by three distinct conductivities in the ridge-parallel (x), ridge-perpendicular (y) and vertical (z) directions. Anisotropy oriented oblique to the ridge was not considered because shear-wave splitting demonstrates that the horizontal symmetry axis for seismic anisotropy is oriented approximately perpendicular to the ridge in the direction of seafloor spreading⁴. The inversion algorithm finds optimally smooth sets of models that fit the data to a desired level of misfit⁵. The degree of anisotropy permitted by the inversion is controlled by a regularization parameter (α) that restricts the level of closeness between the final models in the x , y and z directions. Setting α to 0 allows a freely anisotropic model, while increasing α forces the models in all three directions to be increasingly similar, resulting in an isotropic solution when α is large. All models account primarily for two-dimensional heterogeneity in resistivity structure and anisotropy is only included where needed to decrease misfit further. As expected, the model misfit decreases with decreasing α . Magnetotelluric data are generally not sensitive to the vertical conductivity, except where there are strong, vertical conduction pathways to the surface. Such links do not exist off-axis and so the vertical conductivity structure is not independently resolved. The full results of these inversions are given elsewhere⁶.

Inversions of the magnetotelluric response functions find higher conductivity in the direction of plate spreading with the transition from an electrically isotropic, resistive layer in the shallow upper mantle to an anisotropic conductive region beginning at a depth of around 60 km (Fig. 1). Mantle conductivity is influenced by temperature, composition—including volatile components—and the presence of partial melt. The relatively constant depth of the

resistive-to-conductive boundary at around 60 km indicates that the cause for the change in conductivity is not thermal. This depth is significantly greater than that predicted by models of conductive cooling for the age of the lithosphere in this region. For example, using a standard cooling model⁷ and a potential temperature of 1,350 °C, the depth to the 1,250 °C isotherm increases from only 17 to 29 km between 1.3 and 4.5 Myr ago.

Inversions of Rayleigh- and Love-wave data from the same area also find structure that is nearly independent of age. After a rapid increase in shear velocity to a depth of about 60 km beneath young sea floor within 100 km of the spreading centre, the velocity contours to the east of the ridge are nearly horizontal^{8,9}, decreasing from about 4.5 km s⁻¹ immediately beneath the Mohorovic discontinuity (Moho) to a minimum of about 4.1 km s⁻¹ at about 70 km

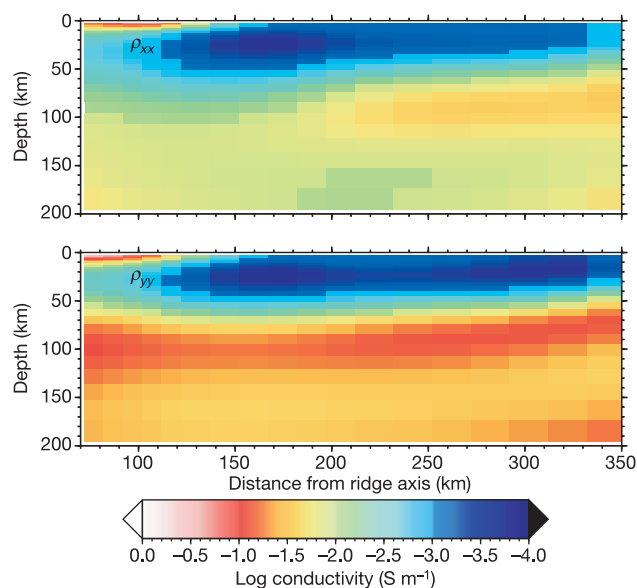


Figure 1 | Two model cross-sections of conductivity through the region 70–350 km east of the southern East Pacific Rise in the MELT area. The two sections show conductivity in the ridge-parallel (x) direction (top) and plate spreading (y) direction (bottom). The model was obtained using an anisotropic regularization parameter, $\alpha = 0.1$, in the inversion. The key features in the model are the resistive layer above about 60 km depth and the underlying conductive region that extends to about 120 km depth. The high-conductivity region is more conductive in the direction of plate spreading, which is also expected to be the direction of dominant a -axis olivine alignment.

¹Department of Geology and Geophysics, ²Department of Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA. ³Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology, 2-15, Natsushima, Yokosuka, Kanagawa, 237-0061, Japan. ⁴Department of Geology, Brown University, USA. ⁵GSY-USA, Inc., PMB#643, 2261 Market Street, San Francisco, California 94114, USA.

(Fig. 2). Here we emphasize structure to the east of the ridge, because the sea floor to the west subsides anomalously slowly, is populated by an unusual density of seamounts, and is poorly constrained by the magnetotelluric coverage. Unlike the conductivity structure, the shear structure in the uppermost 60 km of the mantle is required by the Rayleigh-wave data to be azimuthally anisotropic^{10–12}, with a fast direction perpendicular to the ridge consistent with the fast polarization direction observed from shear-wave splitting analyses⁴. (A discussion of whether electrical anisotropy can be resolved above 60 km is included in the Supplementary Information.) In addition, the magnitude of shear-wave splitting suggests that the anisotropic layer extends much deeper than 60 km.

To compare the results of our inversions to laboratory constraints on the conductivity of mantle minerals we calculated laterally averaged log-conductivity depth profiles from our models for both the plate-spreading (y) and ridge-parallel (x) directions. Examples are shown in Fig. 3 for models with $\alpha = 0.1$ (moderately anisotropic) and $\alpha = 10$ (effectively isotropic). Regardless of the degree of anisotropy in the inversions, a persistent feature is a large region of high conductivity east of the ridge. For the anisotropic models—that is, with $\alpha < 10$ —the conductivity parallel to the spreading direction is always greater than that parallel to the ridge axis. The anisotropy in conductivity is greatest from depths of ~ 70 to 150 km.

In the isotropic model (Fig. 3a) the conductivity at depths greater than ~ 100 km is significantly greater than that predicted from laboratory data for dry peridotite at a potential temperature of $1,350^\circ\text{C}$. The laboratory-based conductivity profiles are calculated using a temperature profile for 3-Myr-old lithosphere, assuming a mantle comprised of 75% olivine and 25% orthopyroxene¹³. There is almost no error associated with extrapolation of the experimental data because the measurements were made at the conditions of interest. However, the laboratory data¹⁴ exhibit a sample-to-sample variability of the order of a factor of two in conductivity. While the conductivity at depths greater than 100 km could be explained by a potential temperature of $\sim 1,500^\circ\text{C}$, this value is considerably greater than typically predicted for the upper mantle beneath the southern East Pacific Rise.

At depths greater than 100 km, the laterally averaged conductivity parallel to the spreading direction in our anisotropic model is similar to that predicted by the Nernst–Einstein relation for hydrogen diffusion parallel to the $[100]$ direction in olivine containing $10^3 \text{ H}/10^6 \text{ Si}$ (Fig. 3b)¹⁵. We emphasize that the plots in Fig. 3 are calculated using hydrogen self-diffusivities¹⁶, which is appropriate for application of the Nernst–Einstein relation. The hydrogen self-diffusivity is estimated from the analysis of hydrogen chemical diffusion data^{16,17}. Even including the uncertainties in the hydrogen diffusion data, calculated conductivities predicted for the other

crystallographic directions are significantly smaller (Fig. 3c). Therefore, anisotropy in conductivity is possible if there is a strong lattice preferred orientation of olivine in the peridotite imposed by corner flow. The factor-of-four anisotropy predicted for the laterally averaged model in Figs 3b and c is the same as that calculated for peridotite with $\sim 50\%$ alignment of the olivine $[100]$ axes¹⁸, a value consistent with the 3%–5% azimuthal anisotropy of Rayleigh waves observed in this region.

An important caveat to the application of the hydrogen diffusion data to constrain mantle conductivity is that the influence of hydrogen on conductivity of olivine has not been measured experimentally, although it has been observed for wadsleyite and ringwoodite¹⁹. A critical factor to be determined is the fraction of hydrogen that contributes to the conductivity^{17,19}. In our calculations we have assumed that all of the hydrogen is contributing, and thus we predict a maximum influence of hydrogen on conductivity.

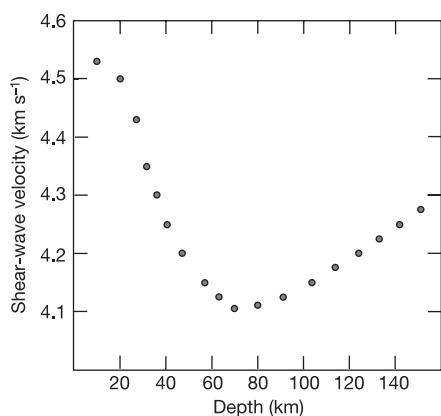


Figure 2 | Shear-wave velocity profile obtained through inversion of Rayleigh-wave data in the same region as Fig. 1.

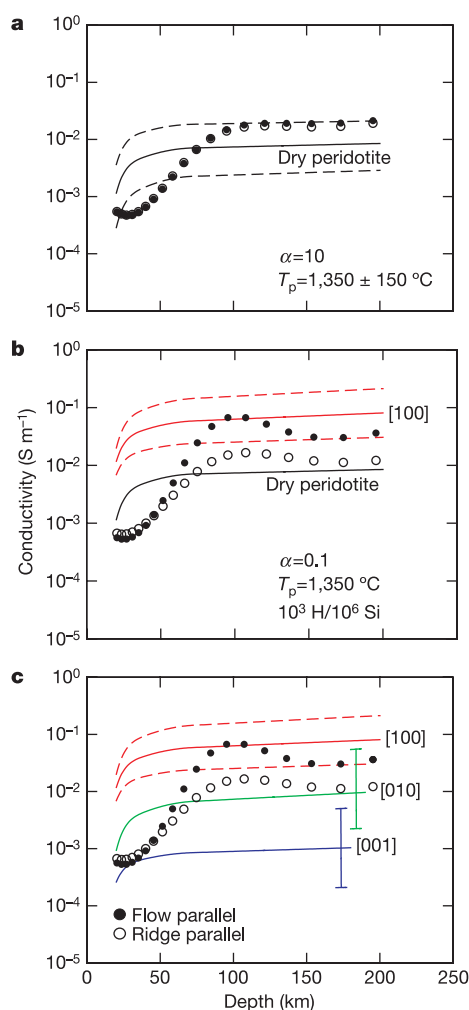


Figure 3 | Average one-dimensional profiles through our conductivity model and theoretical curves based on laboratory data. Comparisons between averaged values of conductivity from the MELT area with laboratory data on mantle conductivity. **a**, An isotropic model compared to conductivity of a dry peridotite mantle on a 3-Myr-old geotherm and an adiabat with a potential temperature of $T_p = 1,350^\circ\text{C}$. The dashed lines show the effects of a variation in adiabat temperature of 150°C . **b**, Average conductivities through an anisotropic model compared to the predicted conductivity along the a axis of olivine shown with estimates of experimental errors. **c**, Same as **b** but with the conductivities in the three crystallographic directions shown as labelled. We used an olivine hydrogen content of $10^3 \text{ H}/10^6 \text{ Si}^3$ and hydrogen diffusion data¹⁶ to calculate conductivity using the Nernst–Einstein relation.

Because dissolved hydrogen influences the concentration of other defects within olivine, conductivity could also be enhanced under hydrous conditions by an increase in the concentration of metal vacancies. However, analyses based on the measured effect of water on Mg–Fe interdiffusion in olivine suggest that this effect is not large enough to explain the high conductivities observed in the oceanic mantle²⁰.

Two key sets of observations lead us to conclude that the mantle above 60 km is dehydrated, but significant dissolved hydrogen is present at greater depths. First, although the oceanic mantle is seismically anisotropic from the Moho to depths exceeding 100 km, significant electrical anisotropy and high conductivities are only resolved at depths exceeding 60 km. If the lattice preferred orientation does not change, the simplest explanation is that the higher anisotropic conductivities are caused by the anisotropic diffusivity of hydrogen. The orientation of the high-conductivity direction coincides with the fast direction for seismic wave propagation, as is expected if both are controlled by the lattice preferred orientation of olivine crystals in a deforming mantle. Second, if thermal effects were dominant, both the shear velocity and resistivity should gradually increase with increasing age and the high-resistivity, high-velocity region should be limited to much shallower depths.

The minimum in shear velocity observed at the same depth as at the start of the increase in conductivity can be caused either by the presence of melt, water (that is, dissolved hydrogen) or a combination of both. Melt affects velocity through elastic and anelastic effects^{21–23}, whereas dissolved water primarily affects velocities through anelastic effects^{24,25}. Therefore, using seismic data to discriminate between the presence of melt and water will require determination of the attenuation structure beneath this region of the mantle. Under the damp conditions we are inferring for the mantle below 60 km, the mantle would be above the solidus, allowing a small amount of melt to be present. The magnetotelluric and seismic data are most consistent with a small melt fraction (<2%) in a hydrous environment at depths greater than 60 km, with a maximum melt concentration at about 70 km. At shallower depths, increased melting and melt migration beneath the ridge axis would extract the water from the mantle minerals and the melt is expected to be removed to form the oceanic crust. The water content remaining in the mantle minerals would be significantly reduced in any residual mantle carried off-axis that had reached melt fractions greater than ~0.5% (refs 3, 26).

An alternative explanation for the abrupt increase in conductivity below 60 km is that the conductivity is controlled by melt below this boundary. In this case the change in physical properties would still represent a compositional boundary because melting at these conditions requires the presence of volatiles. If melt is the explanation for the conductivity enhancement, then the 60-km transition must act as an impermeable boundary, impeding the melt from rising. However, to account for the anisotropic conductivity, the melt would need to be distributed in tubes parallel to the spreading direction—and there is no known mechanism that would cause the melt to reside in such tubes. In addition, the abrupt change from resistive to conductive structure at approximately 60 km is also observed directly beneath the ridge axis where melting is ongoing at depths shallower than 60 km (ref. 6). The observation that conductivity remains significantly greater than that for dry olivine at depths up to ~200 km in the Pacific mantle²⁷ argues for a solid-state conduction mechanism involving dissolved water, because melting at depths greater than about 80 km requires the presence of water, melting at depths greater than about 125 km is difficult even in damp conditions, and the shear-wave velocity increases at depths greater than 80 km, suggesting the progressive elimination of melt.

Melting beneath the ridge dehydrates only the uppermost 60 km of the mantle, consistent with petrological models that suggest there is a

only a very small degree of melting in the presence of water at greater depths, with higher degrees of melting beginning at about 60 km where the geotherm beneath the ridge axis exceeds the dry solidus. The resulting dry region of upper mantle will have a viscosity significantly greater than the underlying asthenosphere, creating a rheological boundary layer³ that is compositionally and not thermally controlled. Our observations are broadly consistent with seismic data from the central Pacific that show a discontinuity (G) at a depth of 68 km, corresponding to the same dry–wet transition²⁸. This boundary corresponds to the base of the layer of low electrical conductivity and high shear-wave velocity observed in the MELT area.

Received 9 March; accepted 29 June 2005.

1. MELT Seismic Team, Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment. *Science* **280**, 1215–1218 (1998).
2. Evans, R. L. *et al.* Asymmetric electrical structure in the mantle beneath the East Pacific Rise at 17°S. *Science* **286**, 756–759 (1999).
3. Hirth, G. & Kohlstedt, D. L. Water in the oceanic upper mantle: implications for rheology, melt extraction and the evolution of the lithosphere. *Earth Planet. Sci. Lett.* **144**, 93–108 (1996).
4. Wolfe, C. J. & Solomon, S. C. Shear-wave splitting and implications for mantle flow beneath the MELT region of the East Pacific Rise. *Science* **280**, 1230–1232 (1998).
5. Rodi, W. & Mackie, R. L. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion. *Geophysics* **66**, 174–187 (2001).
6. Baba, K., Chave, A., Evans, R. L., Hirth, G. & Mackie, R. Mantle dynamics beneath the East Pacific Rise at 17°S: Insights from the MELT EM data. *J. Geophys. Res.* (submitted).
7. Parsons, B. & Sclater, J. Variation of ocean floor bathymetry and heat flow with age. *J. Geophys. Res.* **82**, 803–827 (1977).
8. Conder, J. A., Forsyth, D. W. & Parmentier, E. M. Asthenospheric flow and the asymmetry of the East Pacific Rise, MELT area. *J. Geophys. Res.* **107**(B12), 2344, doi:10.1029/2001JB000807 (2002).
9. Dunn, R. A. & Forsyth, D. W. Imaging the transition between the region of mantle melt generation and the crustal magma chamber with short-period Love wave propagation along the southern East Pacific Rise. *J. Geophys. Res.* **108**(B7), 2352, doi:10.1029/2002JB002217 (2003).
10. Nishimura, C. & Forsyth, D. W. The anisotropic structure of the upper mantle in the Pacific. *Geophys. J. R. Astron. Soc.* **96**, 203–229 (1989).
11. Forsyth, D. W., Webb, S. C., Dorman, L. M. & Shen, Y. Phase velocities of Rayleigh waves in the MELT experiment on the East Pacific Rise. *Science* **280**, 1235–1238 (1998).
12. Forsyth, D. W. & Li, A. Array-analysis of two-dimensional variations in surface wave velocity and azimuthal anisotropy in the presence of multipathing interference. In *Seismic Data Analysis and Imaging with Global and Local Arrays* (eds Levander, A. & Nolet, G.) (AGU Geophys. Monogr., in the press).
13. Xu, Y. S., Shankland, T. J. & Poe, B. T. Laboratory-based electrical conductivity in the Earth's mantle. *J. Geophys. Res.* **105**, 27865–27875 (2000).
14. Xu, Y. S., Shankland, T. J. & Duba, A. G. Pressure effect on electrical conductivity of mantle olivine. *Phys. Earth Planet. Inter.* **118**, 149–161 (2000).
15. Karato, S. The role of hydrogen in the electrical conductivity of the upper mantle. *Nature* **347**, 272–273 (1990).
16. Kohlstedt, D. L. & Mackwell, S. J. Diffusion of hydrogen and intrinsic point defects in olivine. *Z. Phys. Chem.* **207**, 147–162 (1998).
17. Kohlstedt, D. & Mackwell, S. J. In *Microscopic Properties and Processes in Minerals* (eds Wright, K. & Catlow, R.) 539–559 (Kluwer Academic, Dordrecht, 1999).
18. Simpson, F. Intensity and direction of lattice preferred orientation of olivine: are electrical and seismic anisotropies of the Australian mantle reconcilable? *Earth Planet. Sci. Lett.* **203**, 535–547 (2002).
19. Huang, X., Xu, Y. & Karato, S. Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite. *Nature* **434**, 746–749 (2005).
20. Hier-Mujander, S., Anderson, I. M. & Kohlstedt, D. L. Influence of protons on Fe–Mg interdiffusion in olivine. *J. Geophys. Res.* **110**, B02202, doi:10.1029/2004JB003292 (2005).
21. Hammond, W. C. & Humphreys, E. D. Upper mantle seismic wave velocity: Effects of realistic partial melt geometries. *J. Geophys. Res.* **105**(B5), 10975–10986, doi:10.1029/2000JB900041 (2000).
22. Hammond, W. C. & Humphreys, E. D. Upper mantle seismic wave attenuation: Effects of realistic partial melt distribution. *J. Geophys. Res.* **105**(B5), 10987–11000, doi:10.1029/2000JB900042 (2000).
23. Faul, U. H., Fitz Gerald, J. D. & Jackson, I. Shear wave attenuation and dispersion in melt-bearing olivine polycrystals: 2. Microstructural interpretation and seismological implications. *J. Geophys. Res.* **109**, B06202, doi:10.1029/2003JB002407 (2004).
24. Karato, S. Mapping water content in upper mantle. In *Inside the Subduction*

- Factory (ed. Eiler, J.) 135–152 (Geophys. Monogr. Ser. 138, American Geophysical Union, Washington DC, 2003).
25. Karato, S. Effects of water on seismic-wave velocities in the upper-mantle. *Proc. Jpn. Acad. Ser. B*, **71**, 61–66 (1995).
 26. Aubaud, C., Hauri, E. H. & Hirschmann, M. H. Water partition coefficients between nominally anhydrous minerals and basaltic melts. *Geophys. Res. Lett.* (in the press).
 27. Lizarralde, D., Chave, A. D., Hirth, G. & Schultz, A. Northeastern Pacific mantle conductivity profile from long-period magnetotelluric sounding using Hawaii-to-California submarine cable data. *J. Geophys. Res.* **100**, 17837–17854 (1995).
 28. Gaherty, J., Jordan, T. H. & Gee, L. S. Seismic structure of the upper mantle in a central Pacific corridor. *J. Geophys. Res.* **101**, 22291–22309 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank S. Karato for comments. We are also grateful to S. Hier-Mujander and D. Kohlstedt for discussions. US participation in the MELT experiment and subsequent analysis was funded by NSF grants through the Marine Geology and Geophysics Program, Ocean Sciences Division.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to R.L.E. (revans@whoi.edu).

The effect of migration on local adaptation in a coevolving host-parasite system

Andrew D. Morgan¹, Sylvain Gandon² & Angus Buckling¹

Antagonistic coevolution between hosts and parasites in spatially structured populations can result in local adaptation of parasites^{1–5}; that is, the greater infectivity of local parasites than foreign parasites on local hosts¹. Such parasite specialization on local hosts has implications for human health and agriculture. By contrast with classic single-species population-genetic models^{6,7}, theory indicates that parasite migration between subpopulations might increase parasite local adaptation, as long as migration does not completely homogenize populations^{8–11}. To test this hypothesis we developed a system-specific mathematical model and then coevolved replicate populations of the bacterium *Pseudomonas fluorescens* and a parasitic bacteriophage with parasite only, with host only or with no migration. Here we show that patterns of local adaptation have considerable temporal and spatial variation and that, in the absence of migration, parasites tend to be locally maladapted. However, in accord with our model, parasite migration results in parasite local adaptation, but host migration alone has no significant effect.

Conventional wisdom holds that parasites have an evolutionary advantage when antagonistically coevolving with their hosts; they are therefore expected to be locally adapted to their hosts, rather than vice versa². Several studies support this prediction^{12–15}, but others find no evidence of local adaptation, and some even find parasite local maladaptation^{16–19}. Recent theoretical work may be able to explain these varied empirical results. The species that can adapt the fastest is most likely to be locally adapted, and speed of adaptation can be limited by the amount of within-population genetic variation^{8–11}. An important source of genetic variation is migration between populations, because it potentially allows the introduction of genes beneficial at a particular point in space and time^{8–11}. If we assume that migration rates are not so high as to homogenize populations completely, and other factors that might influence local adaptation are equal between hosts and parasites, parasites are predicted to be locally adapted if they migrate more than their hosts, whereas hosts are predicted to be locally adapted if they migrate more than their parasites^{8–11}.

Recent experimental studies have demonstrated that migration can alter coevolutionary dynamics^{20,21}, but whether relative migration rates of hosts and parasites are a significant determinant of local adaptation has yet to be empirically investigated. We examined the role of differential migration rates of hosts and parasites on parasite local adaptation with the use of the aerobic Gram-negative bacterium *Pseudomonas fluorescens* as the host and an associated lytic DNA phage SBW25 ϕ 2 as the parasite¹⁹. Large population sizes and short generation times favour rapid evolution²², and the antagonistic nature of the interaction (phages must lyse the bacteria to release progeny virions) results in sustained coevolution between bacterial resistance and phage infectivity¹⁹, as well as favouring the evolution of local adaptation⁹. Detecting

coevolutionary change is made possible by the ability to store populations in ‘suspended animation’ in a freezer, which allows interactions to be directly measured across space and time¹⁹.

Throughout this study, we quantify parasite local adaptation as the performance of local parasites minus that of the average performance of foreign parasites (‘local versus foreign’)¹, where parasite performance is measured as the proportion of bacterial clones sensitive to phage. A positive value indicates parasite local adaptation and a negative value parasite local maladaptation. Note that parasite local maladaptation does not necessarily imply host local adaptation, when considering a single focal population. However, when considering data averaged across all populations, mean parasite local adaptation equates with mean host local maladaptation (see Supplementary Information).

We performed simulation studies to show that theoretical predictions about the impact of migration on local adaptation hold true using parameter values that approximate our biological system and experimental design (see Supplementary Information for more details). A key parameter in determining coevolutionary dynamics is the specificity of the interaction between host and parasite^{23,24}. Because we do not know the underlying model of specificity governing the interaction between the bacteria and the phage, we used a general model of coevolution that allowed us to explore different circumstances²³ (see Supplementary Methods). At one extreme of this model the different host and parasite genotypes

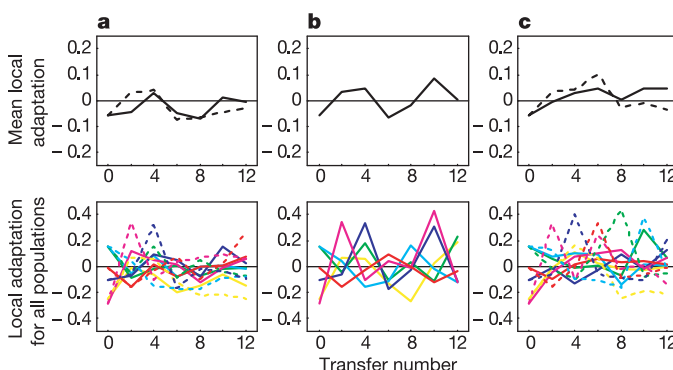


Figure 1 | The simulated effect of differential migration rates on local adaptation. Parasite local adaptation is shown through time for individual populations (bottom) and for the average over the different populations (top). **a**, Host migration only; **b**, no migration; **c**, parasite migration only. In **a** and **c**, solid lines show 10% migration, dashed lines 0.1% migration. Other parameter values are as follows: $\mu_h = \mu_p = 10^{-6}$, $c = k = 0.05$, $\sigma_h = 0.9$, $\sigma_p = 0.5$, $a = 0.5$. See Supplementary Information for model details. A positive value indicates that the parasite is locally adapted, and a negative value that the parasite is locally maladapted.

¹Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ²Génétique et Évolution des Maladies Infectieuses, UMR CNRS/IRD 2724, IRD, 911 avenue Agropolis, 34394 Montpellier Cedex 5, France.

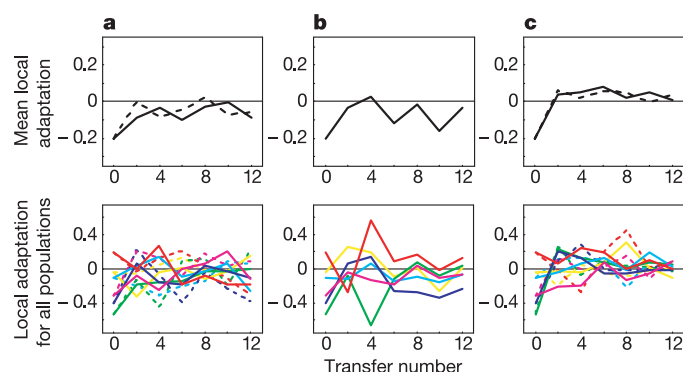


Figure 2 | The effect of differential migration rates on local adaptation. Parasite local adaptation is shown through time for individual populations (bottom) and for the average over the different populations (top). **a**, Host migration only; **b**, no migration; **c**, parasite migration only. In **a** and **c**, solid lines show 10% migration, dashed lines 0.1% migration.

have identical levels of specificity in resistance and infectivity, respectively ('matching alleles' model, when $a = 0$)²³. At the other extreme, the model allows variable levels of specificity among the different host and parasite genotypes, allowing for the evolution of generalists ('gene for gene' model, when $a = 1$)²³. Under a pure 'gene for gene' model the spread of universally resistant hosts and universally infective parasites prevents the emergence of local adaptation⁹. However, our simulations show that coevolution where generalists can evolve (that is, $a > 0$), as occurs in our experimental system¹⁹, can result in local adaptation for a broad range of parameter values (provided that $a < 1$; see Supplementary Fig. S2).

Figure 1 shows an example of a simulation where parasites are locally maladapted in the absence of migration, as occurs in our experimental system¹⁹ (see Figs 2 and 3). Note that host and parasite populations were founded with 10^6 individuals, and measures of local adaptation are calculated from all individuals. The figure illustrates the fact that, when measured at the scale of a single population, antagonistic coevolution yields large fluctuations of local adaptation through time^{3,8,25}. However, when measured at the scale of the metapopulation (that is, when local adaptation is averaged over different populations), the effect of the migration treatment is more apparent. Phage migration increased phage local adaptation, whereas bacterial migration had little effect. Supplementary Fig. S2 gives means and standard deviations obtained after 1,000 runs for a range of parameter values.

To test the role of host and parasite migration on local adaptation experimentally, we first established six replicate populations of isogenic bacteria and phages in experimental microcosms (static glass tubes containing nutrient-rich medium). A 1% portion of culture was transferred (an average of about 2×10^7 bacteria and 10^4 phages) to a fresh microcosm every 2 days, for a total of 73 transfers (about 500 bacterial generations) before experimental treatments, to permit between-population genetic divergence¹⁹. Each of these diverged cultures were used to seed five new cultures (a total of 30 cultures), each of which was exposed to one of five treatments: first, a no-migration control; second, 10% bacteria migration, 0% phage migration; third, 0.1% bacteria migration, 0% phage migration; fourth, 0% bacteria migration, 10% phage migration; and last, 0% bacteria migration, 0.1% phage migration. Cultures were propagated for a further 12 transfers, but bacteria and phages were isolated from each other by chemical means before transfer to a fresh microcosm. In the migration treatments, the specified amount of culture of the migrated species was taken from each replicate within a treatment, added to a common pool and mixed; the same volume was then added back to the replicates. Each migrated replicate was used to inoculate a fresh microcosm along with the previously isolated unmigrated species from the same

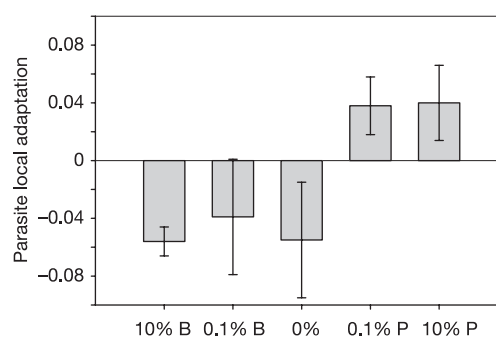


Figure 3 | Mean local adaptation of parasites. Data are averaged over six populations and six time points (T_2 , T_4 , T_6 , T_8 , T_{10} and T_{12}) for different migration treatments. Results are means $\pm$ s.e.m. across populations. P, phage migration; B, bacterial migration.

replicate. If no migration was meant to occur (control populations), isolated bacteria or phages were simply inoculated back together into fresh microcosms. Culture samples were frozen every second transfer. Parasite local adaptation (as defined above) was assayed before the migration treatments (transfer zero), and at six subsequent times (every second transfer). This assay involved measuring the sensitivity of 20 independent bacterial colonies from each population to both sympatric phage populations, and the five allopatric phage populations within the same treatment.

Our experimental data were consistent with our simulations (compare Figs 1 and 2). First, as with other recent experimental data²⁰, we found local adaptation to vary considerably through space and time. Second, phage migration regimes rapidly changed patterns of local adaptation, and resulted in an increase in phage local adaptation, whereas bacterial migration seemed to have little effect. To analyse these data, we calculated mean local adaptation through time (transfers 2, 4, 6, 8, 10 and 12) for each population, and then fitted migration treatment as a five-level factor and local adaptation of starting population (at transfer zero) as a covariate in a General Linear Model. This analysis revealed significant differences in mean phage local adaptation between treatments (Fig. 3; $F_{4,24} = 3.30$, $P = 0.027$), as well as an almost significant effect ($P = 0.06$) of starting local adaptation. Simplification of the statistical model by pooling treatments²⁶ revealed no significant difference between the parasite migration treatments, and no differences between the host migration and the no-migration treatments (Fig. 3; $F_{3,24} = 0.08$, $P > 0.5$). Thus parasite migration, regardless of the rate, increased parasite local adaptation relative to when either host migration or no migration occurred ($F_{1,27} = 14.40$, $P = 0.001$).

It is interesting that the effect of migration was asymmetric: parasite migration increased parasite local adaptation relative to unmigrated populations, but host migration did not decrease parasite adaptation. This is likely to be because parasites are locally maladapted in the absence of any migration, and host migration is unable to increase parasite local maladaptation any further. Such a result was obtained in our theoretical analysis when the strength of selection on the host was higher than on the parasite ($\sigma_h > \sigma_p$): the parasite is expected to be locally maladapted when there is no migration, and the benefits of increased genetic variation resulting from host migration are diluted by stronger selection (Supplementary Fig. S2). An alternative explanation for this asymmetry is that within-population genetic variation for bacterial resistance may be greater than within-population genetic variation for phage infectivity; hence phages are locally maladapted in the absence of migration, and increased bacterial migration provides relatively little additional genetic variation for resistance traits. A difference in within-population genetic variation could result from the larger genome (hence more ways in which to generate resistance)²⁷ and population sizes¹⁹ of bacteria than those of phages.

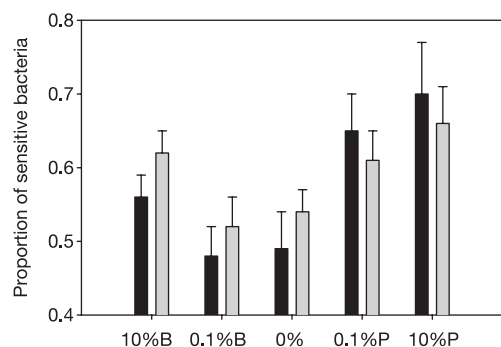


Figure 4 | Mean parasite infectivity. Shown are the proportions of sensitive bacteria to sympatric (black bars) and allopatric (grey bars) phage populations (averaged over six populations and six time points, namely T_2 , T_4 , T_6 , T_8 , T_{10} and T_{12}) for different migration treatments. Results are means $\pm$ s.e.m. across populations. P, phage migration; B, bacterial migration.

We also determined how host and parasite migration affected host local adaptation: the difference in performance between local and foreign hosts. Although the sign-independent mean values of host local adaptation were the same as parasite local adaptation, the variance was much greater, resulting in no significant effect of migration treatment ($F_{4,24} = 1.02$, $P = 0.4$). How can we explain this discrepancy? In this system, bacterial resistance genes that are beneficial against any particular phage populations (local resistance) are likely to be beneficial against all phage populations (global resistance); there is a strong covariance between local and global resistance; see Supplementary Information). Some bacterial populations will therefore be more resistant than other bacterial populations to all phage populations, hence obscuring evidence of host local adaptation. By contrast, the covariance between local and global infectivity of phage was much weaker, favouring the detection of parasite local adaptation.

An increased probability that either host or parasite will be locally adapted through differential migration is believed to result from host or parasite having an evolutionary advantage, and hence being ahead in a coevolutionary arms race^{8–11}. We therefore expected average levels of phage infectivity to sympatric bacteria (defined as 'local performance'; see Supplementary Information) to be higher when phages were migrated than when there was no migration or bacterial migration. Our analysis, which also includes sympatric infectivity before migration treatments as a covariate, support this prediction (Fig. 4; $F_{4,24} = 5.61$, $P = 0.002$). Differences in sympatric infectivity between treatments may reflect two, not mutually exclusive, beneficial effects of migration: first, the introduction of genes that are beneficial for infecting a specific host population (the significant increase in phage local adaptation resulting from phage migration demonstrates that this process occurred); and second, the introduction of genes that are useful for phage infectivity across all populations²⁸ (this possibility was not considered in our theoretical analysis). To investigate the importance of the latter process, we determined how migration rates affected average infectivity on allopatric bacteria, also fitting allopatric infectivity before migration as a covariate. We found a significant effect of migration treatment (Fig. 4; $F_{4,24} = 3.06$, $P = 0.04$), indicating that migration might also allow universally beneficial genes to spread through populations. Note that starting sympatric and allopatric infectivity (at transfer zero) were significant positive predictors of sympatric and allopatric infectivity, respectively, in the above analyses ($P < 0.01$), emphasizing the significance of historical contingency in current infectivity patterns.

We have thus demonstrated experimentally that parasite migration can provide an evolutionary advantage to parasites in a host–parasite coevolutionary arms race, and can therefore result in

parasite local adaptation. That migration can enhance local adaptation is somewhat counterintuitive, given that migration can homogenize populations. However, our data are consistent with theory, which indicates that low rates of migration might increase within-population genetic variation of the migrated species, increasing evolutionary potential. We have yet to address whether local adaptation is influenced by relative migration rates when both species migrate, as predicted by theory^{8–11}. The global scale of human interaction indicates that these results might have important implications for the evolution of human diseases. Parasites will inevitably accompany humans, but parasites will probably remain in the locality and contribute to the local gene pool more often than their hosts. Global interactions may therefore not only increase parasite transmission but may also increase the infectivity of parasites to their local host populations.

METHODS

Generation of divergent populations. Six replicate populations were initiated with approximately 10^8 cells of isogenic *Pseudomonas fluorescens* SBW25 (ref. 29) and about 10^5 particles of phage SBW25 ϕ 2 (ref. 19). Cultures were grown at 28 °C in 6 ml of King's media B (KB) in static 30-ml glass universal flasks with loose plastic lids. Every 48 h, 60 μ l of culture was transferred to a fresh microcosm for a total of 73 transfers, to allow divergence between populations.

Isolation of bacteria and phages. The experiments required bacteria and phage to be isolated from each other. To isolate phage from bacteria, 100 μ l of chloroform was added to 900 μ l of culture, vortex-mixed and then centrifuged at 13,000 r.p.m. for 2 min (ref. 19). This lysed and pelleted the bacterial cells to the bottom of the tubes, leaving a suspension of phages in the supernatant (chloroform is denser than the supernatant and sits at the base of the tube). These were stored at 4 °C until needed. To isolate bacteria, a 5% solution of Virkon (a commercially available disinfectant) in water was made. This was added to KB to a concentration of 0.37% Virkon. A 60- μ l sample of culture was added to 6 ml of the Virkon/KB solution in glass universal flasks, and left for 24 h, static at 28 °C. This procedure left the bacteria viable and completely phage free. A 60- μ l portion of the Virkon cultures was added to a fresh static glass universal flask containing 6 ml of KB, and grown for one day at 28 °C to give a phage-free and Virkon-free stock. Phage-free bacterial cultures were stored in 20% glycerol solution at –80 °C. The potential presence of phages in Virkon-treated cultures was regularly tested for by plating out the cultures on to semisolid agar seeded with *Pseudomonas fluorescens* strain SBW25 and incubated at 28 °C for 24 hr. Plaques would have indicated the presence of phages; in all cases the test proved negative.

Migration regimes. Five treatments were used for this experiment: control; 10% bacteria migration, 0% phage migration; 0.1% bacteria migration, 0% phage migration; 0% bacteria migration, 10% phage migration; and 0% bacteria migration, 0.1% phage migration. A 60- μ l sample of culture from the six diverged cultures was inoculated into five new microcosms per diverged culture (T_0), one for each of the treatments (six replicates per treatment, five treatments, yielding a total of 30 microcosms), and cultured for one transfer (48 h; T_1), after which bacteria or phages were isolated from each culture if they were not to be migrated (as described in the previous section). When phages or bacteria were migrated, 10% or 0.1% (depending on the migration regime) of culture from each of the six populations within a treatment was removed to a common pool, and mixed. The same volume of culture removed from the replicates was removed from the common pool and added back to the replicates. The organism to be migrated was isolated from the culture (as detailed in the previous section) and the rest of the culture was discarded. For each replicate, 60 μ l of the isolated, migrated organism was added to a fresh microcosm along with 60 μ l of the previously isolated unmigrated organism. In unmigrated control replicates, bacteria and phages were isolated from each other at each transfer, using the same chemical treatments as the migrated replicates, then 60 μ l of the isolated bacteria and phages were used to inoculate fresh microcosms. Cultures were frozen every two transfers at –80 °C in 20% glycerol. Phages were isolated every two transfers and stored for local adaptation assays. Populations were evolved for 12 experimental transfers.

Local adaptation assay. We measured local adaptation in terms of bacterial resistance and phage infectivity. Every possible bacteria–phage pairwise interaction within each treatment was measured (36 (that is, 6×6) per treatment). For each interaction, 20 independent bacterial colonies were streaked across a 20 μ l line of phages that had been dried on a KB agar plate¹⁹. Resistance and infectivity were measured as a binary trait, where a bacterial colony was classed as

sensitive if there was any visible inhibition of growth when exposed to the phage, and resistant otherwise. These assays were performed after 2, 4, 6, 8, 10 and 12 transfers, as well as immediately before the migration treatments.

Received 18 February; accepted 13 June 2005.

- Kawecki, T. D. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
- Ebert, D. Virulence and local adaptation of a horizontally transmitted parasite. *Science* **265**, 1084–1086 (1994).
- Kaltz, O. & Shykoff, J. A. Local adaptation in host–parasite systems. *Heredity* **81**, 361–370 (1998).
- Thompson, J. N. The evolution of species interactions. *Science* **284**, 2116–2118 (1999).
- Thompson, J. N. Specific hypotheses on the geographic mosaic of coevolution. *Am. Nat.* **153**, S1–S14 (1999).
- Slatkin, M. Gene flow and the genetic structure of natural populations. *Science* **236**, 787–792 (1987).
- Lenormand, T. Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**, 183–189 (2002).
- Gandon, S., Capowiez, Y., Dubois, Y., Michalakis, Y. & Olivieri, I. Local adaptation and gene for gene coevolution in a metapopulation model. *Proc. R. Soc. Lond. B* **263**, 1003–1009 (1996).
- Lively, C. M. Migration, virulence and the geographic mosaic of adaptation by parasites. *Am. Nat.* **153**, S34–S47 (1999).
- Gandon, S. & Michalakis, Y. Local adaptation, evolutionary potential and host–parasite coevolution: interactions between migration, mutation, population size and generation time. *J. Evol. Biol.* **15**, 451–462 (2002).
- Gandon, S. Local adaptation and the geometry of host–parasite coevolution. *Ecol. Lett.* **5**, 246–256 (2002).
- Parker, M. A. Local population differentiation for compatibility in an annual legume and its host-specific fungal pathogen. *Evolution* **39**, 713–723 (1985).
- Lively, C. M. Adaptation by a parasitic trematode to local populations of its host. *Evolution* **46**, 1663–1671 (1989).
- Lively, C. M. & Dybdahl, M. F. Parasite adaptation to locally common host genotypes. *Nature* **405**, 679–681 (2000).
- McCoy, K. D., Boulonier, T., Schjorring, S. & Michalakis, Y. Local adaptation of the ectoparasite *Ixodes uriae* to its seabird host. *Evol. Ecol. Res.* **4**, 441–456 (2002).
- Kaltz, O., Gandon, S., Michalakis, Y. & Shykoff, J. A. Local maladaptation in the anther-smut fungus *Microbotryum violaceum* to its host *Silene latifolia*: Evidence from a cross inoculation experiment. *Evolution* **53**, 395–407 (1999).
- Mutikainen, P., Salonen, V., Puustinen, S. & Koskela, T. Local adaptation, resistance, and virulence in a hemiparasitic plant–host plant interaction. *Evolution* **54**, 433–440 (2000).
- Oppliger, A., Vernet, R. & Baez, M. Parasite local maladaptation in the Canarian lizard *Gallotia galloti* (Reptilia: Lacertidae) parasitized by haemogregarin blood parasite. *J. Evol. Biol.* **12**, 951–955 (1999).
- Buckling, A. & Rainey, P. B. Antagonistic coevolution between a bacterium and a bacteriophage. *Proc. R. Soc. Lond. B* **269**, 931–936 (2002).
- Forde, S. E., Thompson, J. N. & Bohannan, B. J. M. Adaptation varies through space and time in a coevolving host–parasitoid interaction. *Nature* **431**, 841–844 (2004).
- Brockhurst, M. A., Morgan, A. D., Rainey, P. B. & Buckling, A. Population mixing accelerates coevolution. *Ecol. Lett.* **6**, 975–979 (2003).
- Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2000 generations. *Am. Nat.* **138**, 1315–1341 (1991).
- Agrawal, A. & Lively, C. M. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. *Evol. Ecol. Res.* **4**, 79–90 (2002).
- Frank, S. A. Statistical properties of polymorphism in host–parasite genetics. *Evol. Ecol.* **10**, 307–317 (1996).
- Morand, S., Manning, S. D. & Woolhouse, M. E. J. Parasite–host coevolution and geographic patterns of parasite infectivity and host susceptibility. *Proc. R. Soc. Lond. B* **263**, 119–128 (1996).
- Crawley, M. J. *GLIM for Ecologists* (Blackwell Science, Oxford, 1993).
- Lenski, R. E. Coevolution of bacteria and phage—are there endless cycles of bacterial defenses and phage counterdefenses? *J. Theor. Biol.* **108**, 319–325 (1984).
- Miralles, R., Moya, A. & Elena, S. F. Effect of population patchiness and migration rates on the adaptation and divergence of vesicular stomatitis virus quasispecies populations. *J. Gen. Virol.* **80**, 2051–2059 (1999).
- Rainey, P. B. & Bailey, M. J. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol. Microbiol.* **19**, 521–533 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Brockhurst, T. Day, F. Harrison, O. Kaltz, C. Lively and his group, Y. Michalakis, I. Olivieri, L. Lopez Pascua and A. Rivero for comments on the manuscript. This work was funded by the Royal Society.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.D.M. (andrew.morgan@zoo.ox.ac.uk).

The most infectious prion protein particles

Jay R. Silveira¹, Gregory J. Raymond¹, Andrew G. Hughson¹, Richard E. Race¹, Valerie L. Sim¹, Stanley F. Hayes² & Byron Caughey¹

Neurodegenerative diseases such as Alzheimer's, Parkinson's and the transmissible spongiform encephalopathies (TSEs) are characterized by abnormal protein deposits, often with large amyloid fibrils. However, questions have arisen as to whether such fibrils or smaller subfibrillar oligomers are the prime causes of disease^{1,2}. Abnormal deposits in TSEs are rich in PrP^{res}, a protease-resistant form of the PrP protein with the ability to convert the normal, protease-sensitive form of the protein (PrP^{sen}) into PrP^{res} (ref. 3). TSEs can be transmitted between organisms by an enigmatic agent (prion) that contains PrP^{res} (refs 4 and 5). To evaluate systematically the relationship between infectivity, converting activity and the size of various PrP^{res}-containing aggregates, PrP^{res} was partially disaggregated, fractionated by size and analysed by light scattering and non-denaturing gel electrophoresis. Our analyses revealed that with respect to PrP content, infectivity and converting activity peaked markedly in 17–27-nm (300–600 kDa) particles, whereas these activities were substantially lower in large fibrils and virtually absent in oligomers of ≤ 5 PrP molecules. These results suggest that non-fibrillar particles, with masses equivalent to 14–28 PrP molecules, are the most efficient initiators of TSE disease.

In designing strategies to limit TSE infections and their propagation within hosts, it remains important to identify the most infectious particles and their molecular composition. For several protein aggregation diseases, such as Alzheimer's disease and other amyloidoses, recent studies suggest that instead of being pathological, the formation of large amyloid fibrils might be a protective process that sequesters more dangerous subfibrillar oligomers of the amyloidogenic peptide or protein into relatively innocuous deposits¹. Thus, efforts to disaggregate amyloid deposits might do more harm than good. Although TSE infectivity is associated with a wide range of PrP^{res} aggregate states⁶, little is known about the relative levels of infectivity with respect to particle size, PrP content, or other potential constituents. Furthermore, the size of the smallest infectious unit remains under dispute. Reports that sodium dodecyl sulphate (SDS)-solubilized scrapie PrP monomers can be isolated as infectious units^{7–9} have not been confirmed^{10,11}. Radiation inactivation^{12–14} and liposome solubilization studies¹⁵ have suggested that the minimal infectious unit is between 50–150 kDa, which would correspond to 2–6 PrP molecules, but no such small oligomers have been isolated and determined to be infectious. The present study compares the scrapie infectivity and *in vitro* PrP-converting activity associated with particles ranging in size between the ill-defined smallest infectious unit and large amyloid fibrils, and evaluates which are the most active with respect to PrP content. An infectivity bioassay in hamsters was used to measure infectivity directly, whereas much faster *in vitro* PrP conversion assays were used as surrogates and supplements for the bioassay. Previous studies have shown extensive correlation between infectivity and *in vitro* converting activity^{6,16}.

To break down large PrP^{res} aggregates and create a range of smaller particles for evaluation of these activities, preparations of largely purified PrP^{res} were subjected to treatments with a variety of detergents and sonication. Initial results revealed that relatively small particles with high levels of converting activity were produced when SDS was used at a concentration of $\sim 1\%$ (Supplementary Fig. S1). Subsequent analyses of the alkyl sulphate family of detergents indicated that further optimization was obtained by switching to sodium *n*-undecyl sulphate (SUS) treatment (Supplementary Fig. S2). SUS-treated samples were diluted so that the final SUS concentration (0.1%) was well below the critical micelle concentration, and fractionated according to size using flow field-flow fractionation (FIFFF). Particle molar mass and radius measurements were obtained through in-line light scattering and refractive index detectors. PrP molecules eluted in two major peaks, one spanning fractions 5–8 and another spanning fractions 19–27 (Fig. 1a, filled circles). Dot-blot-based solid-phase conversion assays revealed that converting activity was first observed near fraction 10, rose to a plateau from fractions 12–18, and peaked around fraction 23 (Fig. 1a, open circles). Data obtained from suspension-based conversion assays were consistent with these results, and confirmed that the products observed in the solid-phase conversion assays represented proteinase K-resistant PrP fragments of the expected size³ (Supplementary Fig. S3a). To assess relative infectivity levels, fraction aliquots were also quickly diluted into normal hamster brain homogenate and inoculated intracerebrally into hamsters. A marked shortening of the incubation period of disease was observed near fraction 9, indicating a substantial increase in infectivity¹⁷, which peaked at fraction 12 and remained relatively steady throughout the rest of the elution (Fig. 1a, red squares). Because titration of both purified PrP^{res} and scrapie brain homogenate stocks produced consistent incubation periods of disease with respect to levels of PrP^{res} (Supplementary Fig. S4a), LD₅₀ (dose lethal to 50% of animals tested) values derived from 263K strain scrapie brain homogenates were used to estimate the levels of infectivity in the fractionated material (Supplementary Fig. S4b). This analysis indicated that the 28-day shortening of incubation period between fractions 7 and 12 corresponded to a >600 -fold increase in infectivity. Given limits to the resolution of the FIFFF technique, it is likely that the low levels of infectivity and converting activity detected in fractions 5–8 actually represent the leading edge of activities associated with larger particles that peak in fractions ≥ 9 . These results revealed a major discordance between the levels of PrP, converting activity and infectivity with respect to the size of the particles found in preparations of PrP^{res}.

Because PrP is often thought to be the major protein component of the infectious agent, the levels of converting activity and infectivity in the FIFFF fractions were divided by their total PrP content to give relative 'specific converting activity' (Fig. 1b, circles; see also Supplementary Fig. S3b) and 'specific infectivity' values (Fig. 1b, c, red squares). Both of these values peaked near fraction 12 and dropped

¹Laboratory of Persistent Viral Diseases and ²Electron Microscopy Core Facility, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana 59840, USA.

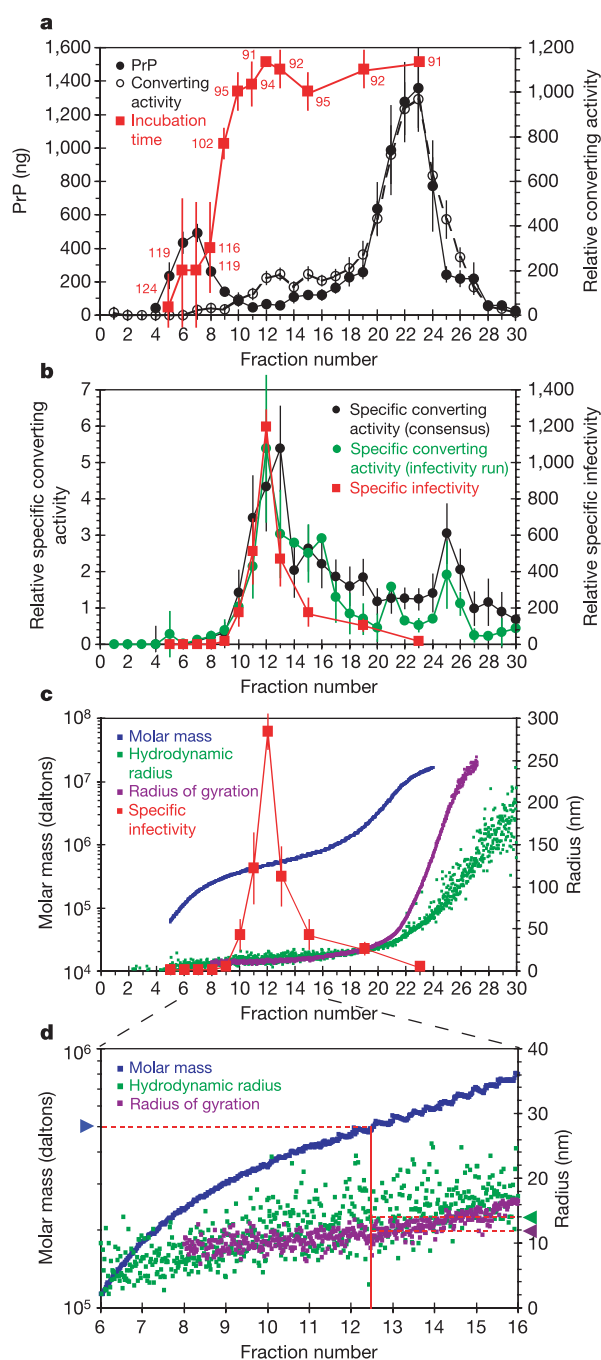


Figure 1 | Analysis of fractionated PrP^{res}. **a**, Quantification of PrP (per 1-ml fractions) by dot immunoblotting with anti-PrP monoclonal antibody 3F4 (n ranged from 3 to 16 from four FIFFF runs), analysis of converting activity by solid-phase conversion assay ($n = 9$ from four FIFFF runs) and scrapie infectivity analysis by incubation time assays in hamsters ($n = 4$ animals from one FIFFF run). Mean incubation periods (days) are indicated by red numbers in the plot. **b**, Calculated specific converting activity for the individual FIFFF run assessed for infectivity ($n = 4$ converting activity analyses), the consensus of all four FIFFF runs ($n = 9$ converting activity analyses) and specific infectivity ($n = 4$ animals). **c**, In-line light scattering analyses of FIFFF runs indicating M_w ($n = 3$), r_h ($n = 4$) and r_g ($n = 4$). The specific infectivity trace from **b** is superimposed. **d**, Expansion of a region of the plot shown in **c**. The solid red line indicates the centre of fraction 12, and the dashed red lines lead to arrowheads, which indicate the mass and radii of particles at this point in the elution. Mean values are shown; error bars represent $\pm$ standard error except for incubation time of disease, where they represent $\pm$ s.d.

off steeply to each side of the peak. Indeed, the early PrP-containing fraction 7 had $\sim 3,000$ -fold lower specific infectivity than the peak fraction 12, and the large aggregates in fraction 23 had ~ 70 -fold lower specific infectivity. This analysis revealed a tight correlation between specific infectivity and specific converting activity, and demonstrated that by far the most active particles per unit PrP peaked in fraction 12. Light scattering analyses of these particles indicated that they have apparent weight-average molar masses (M_w) of several hundred kDa (Fig. 1c, d, blue squares) and apparent radii in the 12–14-nm range (Fig. 1c, d, green and violet squares).

Table 1 summarizes the size parameters determined from four independent FIFFF experiments assessed for converting activity, and

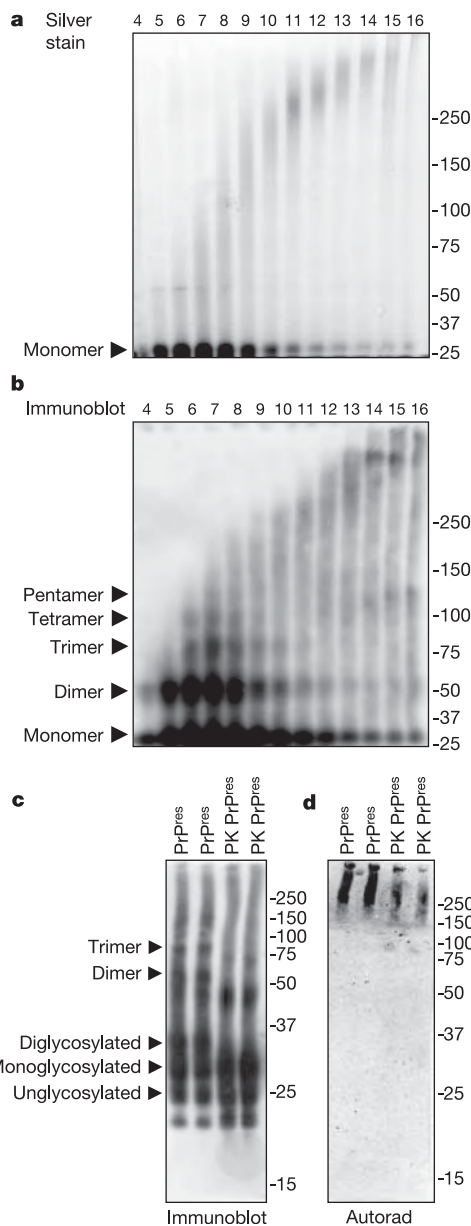


Figure 2 | PAGE analyses of detergent-treated PrP^{res}. **a**, **b**, Samples from FIFFF fractions were subjected to PAGE on 3–8% Tris-acetate gels and analysed by either silver stain (**a**) or immunoblotting with anti-PrP monoclonal antibody 3F4 after transfer to PVDF (**b**). **c**, **d**, Additional samples of purified PrP^{res} were boiled in SDS–PAGE buffer without sonication, subjected to PAGE on 10% Bis-Tris gels, and analysed in duplicate by either immunoblotting (**c**) or solid-phase conversion (**d**). Fraction numbers and types of PrP^{res} used (PK PrP^{res}, proteinase K-digested PrP^{res}) are shown at the top. PrP glycoforms and oligomers are indicated on the left, and molecular weight standards (kDa) are shown on the right.

a single fractionation analysed by infectivity bioassays. The ratio (ρ) of the radius of gyration (r_g) to hydrodynamic radius (r_h) was calculated to estimate the compactness of the particles with peak specific infectivity and converting activity (fraction 12). The values (~ 0.90) are typical of fairly compact, spherical or ellipsoid shapes¹⁸. By comparison, much higher ρ values were obtained for fractions 21 and 26, indicating the predominance of highly extended structures. Mean apparent values of ~ 600 kDa (M_W) and ~ 13.5 nm (r_h) were determined for the material in fraction 12. These values presumably represent the size of the infectious particles including any bound SUS molecules. Analyses of a set of proteins (bovine serum albumin, ferritin, thyroglobulin) in the presence or absence of SUS showed that observed M_W and r_h values could be increased by as much as 73% and 60% respectively in the presence of SUS (data not shown), suggesting that the particles in fraction 12 may have detergent-subtracted M_W values as low as ~ 300 kDa and r_h values as low as ~ 8.5 nm. Thus, these data indicated that the particles with the highest specific infectivity and specific converting activity were approximately 300–600 kDa, roughly spherical-to-elliptical, and 17–27 nm in diameter.

To assess more accurately the oligomeric state of PrP in the early fractions (<9) of the FIFFF separation, samples were subjected to PAGE without further denaturation. Protein (silver) staining (Fig. 2a) and PrP immunoblot (Fig. 2b) analyses of these gels revealed that PrP monomers and a ladder of discrete oligomers up to apparent pentamers (most visible in the immunoblot) peaked in fractions 5–8, which were extremely low in specific infectivity and specific converting activity (Fig. 1b). However, in fractions 11–13, where the highest specific infectivity and specific converting activity were found, monomeric-to-pentameric structures were greatly reduced and larger oligomers above the 250 kDa marker predominated (Fig. 2a, b). The converting activity associated with these oligomers was directly assessed by subjecting detergent-treated PrP^{res} to PAGE as described above, and then electro-blotting the material onto a polyvinylidene difluoride (PVDF) membrane for solid phase conversion analysis. Conversion products were only observed at the bands above the 150–250 kDa markers after SDS (Fig. 2d) or SUS (Supplementary Fig. S5c) treatments, even though the PrP content in the monomer and small oligomer bands was equivalent to, or greatly exceeded, the protein content in the higher bands (Fig. 2c; see also Supplementary Fig. S5a, b). Thus, results obtained from PAGE-based analyses were in excellent agreement with those obtained from dot-blot-based solid-phase conversion analyses (Fig. 1a), and revealed that infectivity and converting activity co-fractionated with structures larger than PrP pentamers, whereas PrP pentamers and smaller oligomers were devoid of converting activity.

To visualize the sizes and shapes of fractionated PrP^{res} particles, samples from representative FIFFF fractions were analysed by transmission electron microscopy (Fig. 3). Consistent with the high ρ values noted in Table 1, fraction 26 contained a preponderance of long fibrils, whereas fraction 21 contained shorter fibrils in conjunction with more amorphous material. Analysis of material from fractions 10 and 15 revealed a collection of smaller amorphous and spherical particles. Although it was unclear whether all of the

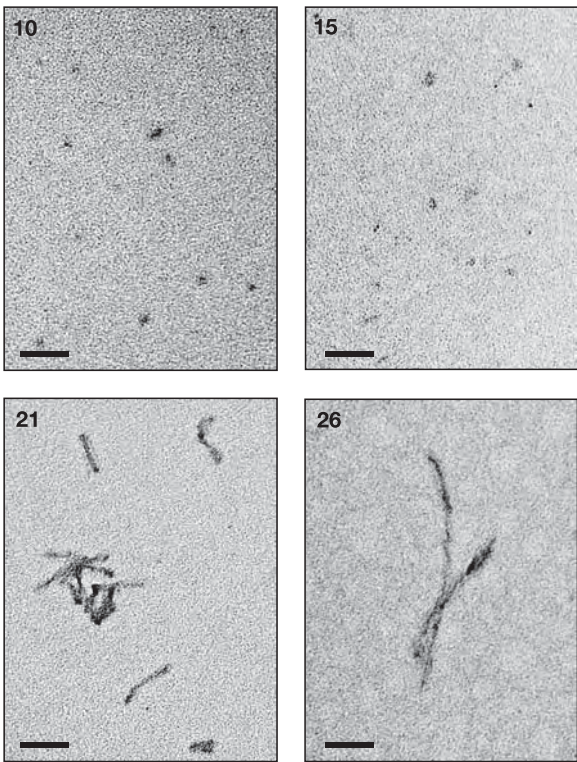


Figure 3 | Transmission electron microscopy analyses of fractionated PrP^{res}. Fraction numbers are indicated in the upper left of each panel. No particles were visible on control grids loaded with buffer alone. Scale bars, 100 nm.

amorphous/spherical particles contained PrP, the results confirmed that there were no visible fibrils in these fractions. Thus, the size and shape of the particles detected by transmission electron microscopy agreed with the light scattering measurements, indicating that the most infectious particles of PrP^{res} were roughly spherical or ellipsoidal in nature, and ~ 20 – 25 nm in diameter.

TSE infectious units (prions) are likely to require both biochemical activity as initiators of PrP conversion and stability against degradation in the environment and the host. Whereas large PrP^{res} aggregates might be expected to have greater stability, smaller oligomers (4–15-mers) have been predicted to have greater converting activity and infectivity per unit mass^{19,20}. Our present data provide systematic evidence that although the infectivity per particle did not vary by more than approximately twofold between different sizes of infectious aggregates (see Supplementary Discussion), the most infectious units per mass of PrP, or the best apparent compromise between stability and activity, are ~ 17 – 27 -nm particles of ~ 300 – 600 kDa. If these infectious particles are composed solely of PrP molecules averaging ~ 21.5 kDa each (which may not be the case, see Supplementary Discussion), this would correspond to an

Table 1 | Biophysical parameters of fractionated PrP^{res}

Fraction	M_W (kDa)	$\langle r_g \rangle_z$ (nm)	$\langle r_h \rangle_z$ (nm)	ρ
6 (monomer/small PrP oligomers)	155 ± 62	ND	5.0 ± 1.7	ND
12 (peak specific infectivity)	535	12.4	13.4	0.93
12 (peak specific converting activity)	620 ± 331	12.1 ± 2.0	13.5 ± 0.9	0.90
21 (intermediate fibrils)	7,770 ± 4,270	51.9 ± 11.5	37.4 ± 6.7	1.38
26 (largest fibrils)	15,220 ± 9,210	230.1 ± 46.3	90.4 ± 8.1	2.35

Weight-average molar mass (M_W), z-average radius of gyration ($\langle r_g \rangle_z$) and z-average hydrodynamic radius ($\langle r_h \rangle_z$) values (mean ± s.d.) were determined from four individual fractionations of PrP^{res}. Values for one fractionation of PrP^{res} assayed for infectivity are also shown. $\rho = r_g/r_h$. ND, not determined.

oligomer of 14–28 PrP molecules. Interestingly, this size range is consistent with the smallest disease-associated PrP aggregates (600 kDa) observed previously²¹. Our observations that the smallest stable unit with PrP converting activity is larger than a 5-mer concur with the results of a previous study that showed that SDS-generated oligomers comprising 4–6 PrP molecules with diameters of ~10 nm are not infectious²². Furthermore, we join others in failing to confirm previous reports of the generation of infectious PrP monomers in the presence of alkyl sulphate detergents^{7–11}. However, although we screened deliberately for conditions that generate small filterable particles with converting activity, it remains possible that other conditions can stabilize infectious particles that are smaller than PrP hexamers.

The fact that the most infectious units are much smaller than the amyloid fibrils that are often observed in TSE-infected tissues and tissue extracts reinforces concerns that incomplete attempts to destabilize PrP^{res} aggregates for the purposes of therapeutics or decontamination might result in unintended increases in infectivity. Consistent with this possibility, we observed that resonication (in 1% SUS) of large fibrillar PrP^{res} fractions from the FIFFF separation decreased the average fibril length (according to electron microscopy) and increased converting activity by several-fold (data not shown), and others have reported that sonication of PrP^{res} in the presence of phospholipids can increase scrapie infectivity levels¹⁵. This is not to say that the most infectious particles are necessarily derived from *in vitro* fragmentation of PrP^{res} fibrils. Another strong possibility is that these particles are derived primarily from a distinct non-fibrillar PrP ultrastructure in TSE-infected brain tissue, such as the commonly observed amorphous membrane-associated deposits²³. Collectively, our observations support the emerging view that with protein folding/aggregation diseases, smaller subfibrillar particles may be much more pathological than larger amyloid fibrils or plaques.

METHODS

Partial disaggregation of PrP^{res}. PrP^{res} was purified from scrapie-infected (263K strain) hamster brain and treated with proteinase K as described previously²⁴ to produce a product of >90% purity. Unless designated otherwise, the samples were pelleted (20,800g, 20 min, 4 °C), re-suspended to 0.1 mg ml⁻¹ in 20 mM Tris pH 7.0 containing 1% SUS, sonicated for 1 min in a cup horn at maximum power, frozen in a dry ice/ethanol bath for 5 min, thawed, and incubated at 37 °C for 1 h. Additional FIFFF experiments in which the sonication step was omitted have shown that the intermediate-sized non-fibrillar oligomers with the highest specific converting activity are still generated, and the recovery of converting activity in fractions containing predominantly PrP monomers to pentamers is not enhanced (data not shown). In other designated experiments, samples of PrP^{res} (0.25 mg ml⁻¹) were simply boiled for 2 min in SDS–PAGE buffer (250 mM Tris pH 7.0, 1% SDS, 5% glycerol, 2.5% β-mercaptoethanol, 0.00025% bromophenol blue) without sonication.

Fractionation of PrP^{res}. Samples of partially disaggregated PrP^{res} (50 µg) were filtered through 0.2 µm Nanosep centrifugal devices (Pall Life Sciences) by centrifugation (500g, 20 min, 25 °C), and the filtrate was subjected to asymmetrical flow field-flow fractionation²⁵ on an Eclipse F separation system (Wyatt Technology Europe). The channel was 26.5 cm in length and 350 µm in height, constructed with a trapezoidal spacer of maximal width 21 mm at the inlet, and lined with a 10-kDa cutoff polyethersulphone membrane at the accumulation wall. The sample was loaded in five 100-µl injections, then eluted with 20 mM Tris pH 7.0 containing 0.1% SUS at a channel flow of 1 ml min⁻¹ and a cross flow decreasing from 3 ml min⁻¹ to 0 ml min⁻¹ over 20 min while collecting 1-ml fractions.

Light scattering analyses. Static light scattering, refractive index and dynamic light scattering measurements were carried out on DAWN EOS, Optilab DSP and WyattQELS instruments, respectively (Wyatt Technologies), connected in-line to the FIFFF system. Weight-average molar mass (M_w), z-average radius of gyration ($(r_g)_z$) and z-average hydrodynamic radius ($(r_h)_z$) values were calculated using ASTRA analysis software (version 4.90.07).

Solid-phase PrP conversion. Samples (50–250 µl) of fractionated PrP^{res} were loaded onto PVDF membranes using a 96-well dot-blot apparatus as described previously²⁶ or transferred by electro-blotting. Blocking and conversion on membrane sheets (~55 cm²) were carried out as described²⁷ under Gdn-free/detergent-free conditions with the following exceptions: conversion reactions

(5 ml) on each membrane sheet contained 500,000 counts per min ³⁵S-labelled PrP^{sen}, and 0.1% fetal bovine serum was replaced with 2% BSA. Proteinase K digestion (5 ml; 10 µg ml⁻¹) was terminated by drying the membrane, which was subjected to autoradiography and PhosphorImager analysis. Relative amounts of conversion in dot-blot analyses were obtained by comparison to a series of PrP^{res} standards (1–1,000 ng) loaded onto each membrane sheet.

Incubation time bioassays for scrapie infectivity. Immediately after FIFFF separation, samples of PrP^{res} were diluted fivefold into physiological phosphate buffer containing 1.25% normal hamster brain homogenate and 40 mM sucrose, and 50 µl aliquots were inoculated intracerebrally into Syrian golden hamsters. The incubation period of disease was defined as the number of days from inoculation to euthanization when confirmed to have clinical disease (recumbent animals).

Transmission electron microscopy. Samples of fractionated PrP^{res} (~5 µl) were placed on 400-mesh parlodian-coated nickel grids, incubated for 1 h, washed with H₂O (three times for 10 min), stained with 1% ammonium molybdate for 30 s, dried and visualized on an 80-kV Hitachi H7500 microscope fitted for image capture with a Hamamatsu side-mounted model C4742-95 CCD camera and Advantage HR/HR-B digital image software (AMT). All steps were performed at ambient temperature, and all reagents were passed through 0.02-µm filters within an hour of use.

Received 15 May; accepted 24 June 2005.

1. Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298 (2003).
2. Kaye, R. et al. Permeabilization of lipid bilayers is a common conformation-dependent activity of soluble amyloid oligomers in protein misfolding diseases. *J. Biol. Chem.* **279**, 46363–46366 (2004).
3. Kocisko, D. A. et al. Cell-free formation of protease-resistant prion protein. *Nature* **370**, 471–474 (1994).
4. Prusiner, S. B. Prions. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
5. Silveira, J. R., Caughey, B. & Baron, G. S. Prion protein and the molecular features of transmissible spongiform encephalopathy agents. *Curr. Top. Microbiol. Immunol.* **284**, 1–50 (2004).
6. Caughey, B., Raymond, G. J., Kocisko, D. A. & Lansbury, P. T. Jr Scrapie infectivity correlates with converting activity, protease resistance, and aggregation of scrapie-associated prion protein in guanidine denaturation studies. *J. Virol.* **71**, 4107–4110 (1997).
7. Prusiner, S. B. Molecular biology of prion diseases. *Science* **252**, 1515–1522 (1991).
8. Brown, P., Liberski, P. P., Wolff, A. & Gajdusek, D. C. Conservation of infectivity in purified fibrillary extracts of scrapie-infected hamster brain after sequential enzymatic digestion or polyacrylamide gel electrophoresis. *Proc. Natl Acad. Sci. USA* **87**, 7240–7244 (1990).
9. Safar, J. et al. Molecular mass, biochemical composition, and physicochemical behaviour of the infectious form of the scrapie precursor protein monomer. *Proc. Natl Acad. Sci. USA* **87**, 6373–6377 (1990).
10. Hope, J. The nature of the scrapie agent: the evolution of the virino. *Ann. NY Acad. Sci.* **724**, 282–289 (1994).
11. Morillas, M., Vanik, D. L. & Surewicz, W. K. On the mechanism of α-helix to β-sheet transition in the recombinant prion protein. *Biochemistry* **40**, 6982–6987 (2001).
12. Alper, T., Haig, D. A. & Clarke, M. C. The exceptionally small size of the scrapie agent. *Biochem. Biophys. Res. Commun.* **22**, 278–284 (1966).
13. Gibbs, C. J. Jr, Gajdusek, D. C. & Lattarjet, R. Unusual resistance to ionizing radiation of the viruses of kuru, Creutzfeldt-Jakob disease, and scrapie. *Proc. Natl Acad. Sci. USA* **75**, 6268–6270 (1978).
14. Bellinger-Kawahara, C. G., Kempner, E., Groth, D., Gabizon, R. & Prusiner, S. B. Scrapie prion liposomes and rods exhibit target sizes of 55,000 Da. *Virology* **164**, 537–541 (1988).
15. Gabizon, R., McKinley, M. P. & Prusiner, S. B. Purified prion proteins and scrapie infectivity copartition into liposomes. *Proc. Natl Acad. Sci. USA* **84**, 4017–4021 (1987).
16. Raymond, G. J. et al. Molecular assessment of the potential transmissibilities of BSE and scrapie to humans. *Nature* **388**, 285–288 (1997).
17. Prusiner, S. B. et al. Molecular properties, partial purification, and assay by incubation period measurements of the hamster scrapie agent. *Biochemistry* **19**, 4883–4891 (1980).
18. Gast, K., et al. in *Laser Light Scattering in Biochemistry* (eds Harding, S. E., Sattelle, D. B. & Bloomfield, V. A.) 209–224 (The Royal Society of Chemistry, Cambridge, 1992).
19. Masel, J., Jansen, V. A. & Nowak, M. A. Quantifying the kinetic parameters of prion replication. *Biophys. Chem.* **77**, 139–152 (1999).
20. Masel, J., Genoud, N. & Aguzzi, A. Efficient inhibition of prion replication by PrP-FC₂ suggests that the prion is a PrP^{Sc} oligomer. *J. Mol. Biol.* **345**, 1243–1251 (2005).
21. Tzaban, S. et al. Protease-sensitive scrapie prion protein in aggregates of heterogeneous sizes. *Biochemistry* **41**, 12868–12875 (2002).

22. Riesner, D. *et al.* Disruption of prion rods generates 10-nm spherical particles having high α -helical content and lacking scrapie infectivity. *J. Virol.* **70**, 1714–1722 (1996).
23. Jeffrey, M. *et al.* Correlative light and electron microscopy studies of PrP localisation in 87V scrapie. *Brain Res.* **656**, 329–343 (1994).
24. Raymond, G. J. & Chabry, J. in *Techniques in Prion Research* (eds Lehmann, S. & Grassi, J.) 16–26 (Birkhauser, Basel, 2004).
25. Wahlund, K.-G. in *Field-Flow Fractionation Handbook* (eds Schimpf, M., Caldwell, K. & Giddings, J. C.) 279–294 (John Wiley & Sons, New York, 2000).
26. Kocisko, D. A. *et al.* New inhibitors of scrapie-associated prion protein formation in a library of 2000 drugs and natural products. *J. Virol.* **77**, 10288–10294 (2003).
27. Maxson, L., Wong, C., Herrmann, L. M., Caughey, B. & Baron, G. S. A solid-phase assay for identification of modulators of prion protein interactions. *Anal. Biochem.* **323**, 54–64 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Y. Huang and D. Follmann (Biostatistics Research Branch, NIH/NIAID) for performing statistical analyses. We thank B. Chesebro, G. S. Baron and S. J. Robertson for critiquing the manuscript. This research was supported in part by the Intramural Research Program of the NIH/NIAID. V.L.S. acknowledges support from the Alberta Heritage Foundation for Medical Research through a clinical fellowship award.

Author Contributions J.R.S. spearheaded the project, developed the critical methods and performed the PrP^{res} disaggregation, fractionation and particle analyses. G.J.R. and A.G.H. purified PrP^{res} and performed bioassays and other supporting experiments. R.E.R. provided bioassay standard curve data. V.L.S. and S.F.H. performed electron microscopy. B.C. helped with project design, data interpretation and writing (primarily with J.R.S.)

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to B.C. (bcaughey@nih.gov).

LETTERS

Prion protein remodelling confers an immediate phenotypic switch

Prasanna Satpute-Krishnan¹ & Tricia R. Serio¹

In a variety of systems, proteins have been linked to processes historically limited to nucleic acids, such as infectivity and inheritance^{1,2}. These atypical proteins, termed prions³, lack sequence homology but are collectively defined by their capacity to adopt multiple physical and therefore functional states *in vivo*. Newly synthesized prion protein generally adopts the form already present in the cell, and this *in vivo* folding bias directs the near faithful transmission of the corresponding phenotypic state^{1,2}. Switches between the prion and non-prion phenotypes can occur *in vivo*²; however, the fate of existing protein during these transitions and its effects on the emergence of new traits remain major unanswered questions. Here, we determine the changes in protein-state that induce phenotypic switching for the yeast prion Sup35/[PSI⁺]. We show that the prion form does not need to be specified by an alternate misfolding pathway initiated during Sup35 synthesis but instead can be accessed by mature protein. This remodelling of protein from one stable form to another is accompanied by the loss of Sup35 activity, evoking a rapid change in cellular phenotype within a single cell cycle.

The [PSI⁺] prion is an alternate, self-perpetuating physical state of the release factor Sup35, leading to decreased translation termination fidelity⁴. Sup35 protein is soluble and stimulates termination in non-prion [*psi*⁻] cells but largely self-associates into functionally compromised complexes in [PSI⁺] cells⁵⁻⁷. During non-prion to prion phenotypic transitions, Sup35^[PSI⁺] is introduced into a cellular pool of soluble and active Sup35^[psi⁻], and colonies derived from these cells ultimately display the prion physical and functional states.

How does the prion phenotype emerge from a cell containing a mixture of Sup35 functional forms? The fate of existing Sup35^[psi⁻] is the crucial variable. For example, if Sup35 is artificially blocked from adopting the prion form in [PSI⁺] cells by the addition of an amino-terminal glutathione S-transferase (GST) tag⁸ (Fig. 1a–c; see also Supplementary Fig. 1a), this unconverted GST–Sup35^[psi⁻] reverses the prion phenotype without disrupting prion propagation by untagged Sup35 (Fig. 1a–c). The prion phenotype only emerges after new GST–Sup35 synthesis ceases (Fig. 1c) and existing protein is diluted by cell division (Fig. 3c).

What, then, becomes of the existing Sup35^[psi⁻] in cells undergoing a [*psi*⁻] to [PSI⁺] switch? To date, this key question has not been resolved by *in vitro* studies, which model the self-perpetuation of protein physical states by assembly of unstructured Sup35 fragments alone or in combination with full-length protein in the presence of denaturant⁹⁻¹¹. One can envision three possible scenarios for the Sup35 protein *in vivo*: degradation, dilution by cell division or remodelling to a new form, the last of which is predicted by the prion hypothesis^{3,12}. To ascertain the fate of Sup35^[psi⁻], we first determined the metabolic stabilities of Sup35^[PSI⁺] and Sup35^[psi⁻]. These conformers are equally stable and long-lived (see Supplementary Fig. 2); thus, existing Sup35 is likely to persist during non-prion to prion transitions.

To determine whether fully synthesized Sup35 can access the prion form, we inhibited Sup35 conversion in [PSI⁺] cells by placing a GST tag at the N terminus, removed this block by proteolytic cleavage, and then followed the fate of the released Sup35 (Fig. 1d). We targeted the ubiquitin (Ub)-like protease Ulp1 or, as a control, the deubiquitinating enzymes to GST–Sup35 by constructing fusion proteins containing their respective recognition tags, the Ub-like protein Smt3 or ubiquitin^{13,14} (Ub, Fig. 1d).

Haploid strains expressing GST–Ub–Sup35 or GST–Smt3–Sup35 as the sole source of Sup35 were viable. As expected¹⁵, GST–Ub was cleaved from the fusion protein during synthesis, as only processed Sup35 was detected in metabolic labelling experiments (see Supplementary Fig. 3a). In contrast, unprocessed GST–Smt3–Sup35 comprised ~50% of the nascent pool in the *GSTSMT3SUP35* strain (Fig. 1e; see also Supplementary Fig. 3a). This tag was subsequently removed with a half-time of ~20 min, leading to a concomitant increase in mature Sup35 levels (Fig. 1e). Although this processing rate does not lead to significant steady-state accumulation of unprocessed protein (see Supplementary Fig. 3b), the transient presence of the GST–Smt3 tag blocked nearly half of the Sup35 in the *GSTSMT3SUP35* strain from adopting the prion form during its synthesis.

Sup35 released from either fusion supported [PSI⁺] propagation *in vivo* (Fig. 1f). If fully synthesized protein could not adopt the prion form after the GST–Smt3 tag was removed, it would accumulate in a soluble and functional form. Indeed, if fusion protein processing is blocked either in *cis* or in *trans*^{13,16}, the prion phenotype is dominantly reversed (Fig. 1a, f) and prion propagation ceases (see Supplementary Fig. 1). In contrast, the [PSI⁺] *GSTSMT3SUP35* strain is phenotypically indistinguishable from its wild type counterpart (Fig. 1f), suggesting that the released Sup35 can adopt the prion form. Consistent with this interpretation, only aggregated Sup35 is detected in [PSI⁺] *GSTSMT3SUP35* lysates (see Supplementary Fig. 3c). Taken together, these results strongly indicate that blocking Sup35^[PSI⁺] formation during translation does not preclude the subsequent conversion of mature Sup35.

Although these experiments reveal that fully synthesized Sup35 can adopt the prion form, they do not directly probe the physical state of the protein species that is competent to convert. To determine if Sup35 could be transformed from one stable state to another, we visualized the remodelling of existing Sup35^[psi⁻] in cells on introduction of Sup35^[PSI⁺] by mating. For these experiments, we expressed a Sup35/green fluorescent protein (GFP) fusion (Sup35–GFP) from *P_{MEAL}*, a promoter that is rapidly repressed on mating¹⁷ (see Supplementary Fig. 4), in strains that constitutively express untagged Sup35 from its endogenous locus. Sup35–GFP is fully functional, acts independently as a prion, and reports on the Sup35 physical state by forming fluorescent foci in the [PSI⁺] cytoplasm and exhibiting diffuse fluorescence in [*psi*⁻] cells (see Supplementary Fig. 5).

¹Department of Molecular Biology, Cell Biology and Biochemistry, Brown University, Providence, Rhode Island 02912, USA.

We mated $[psi^-]$ $MATa$ $P_{MFA1}SUP35-GFP$ haploids to either $[PSI^+]$ or $[psi^-]$ $MAT\alpha$ haploids expressing a red fluorescent protein, DsRed, and identified zygotes by their cloverleaf morphology and coincident red and green fluorescence. Sup35-GFP $^{[psi^-]}$ formed mobile, fluorescent foci in $[PSI^+] \times [psi^-]$ zygotes (Fig. 2a; see also Supplementary Video 1) but remained diffuse in $[psi^-] \times [psi^-]$ zygotes (Fig. 2b; see also Supplementary Video 2).

Fluorescent foci were formed from remodelling of existing Sup35–

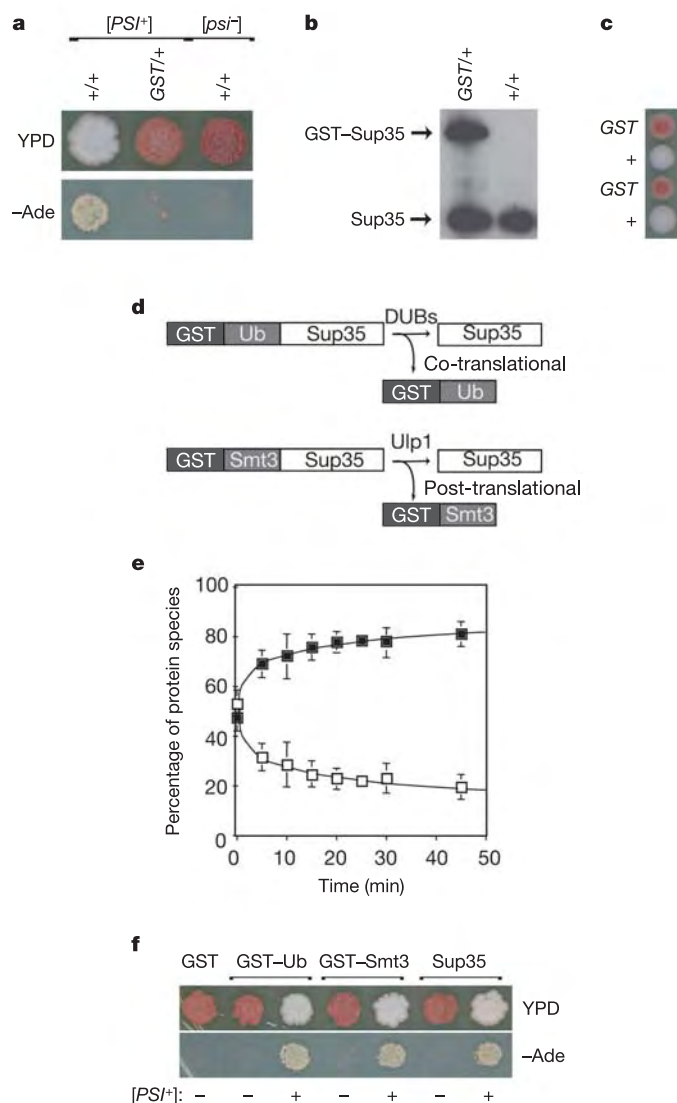


Figure 1 | Mature Sup35 can adopt the prion state. **a**, $[psi^-]$ *ade1-14* (*UGA*) strains require adenine and form red colonies on rich media (YPD), whereas $[PSI^+]$ variants grow without adenine (–Ade) and form white colonies on YPD. GST–Sup35 cannot acquire the prion form and dominantly reverses the $[PSI^+]$ phenotype in a *GSTSUP35/SUP35* $[PSI^+]$ heterozygote (*GST/+*). **b**, Anti-Sup35 immunoblot of lysates from $[PSI^+]$ strains in **a**. **c**, The segregation of two wild type $[PSI^+]$ (+) meiotic progeny indicates that $[PSI^+]$ propagation continued in the *GST/+* diploid through the conversion of wild-type protein despite its phenotype. **d**, Schematic of fusion protein processing. DUB, deubiquitinating enzymes; Ulp1, ubiquitin-like protease 1; Ub, ubiquitin; GST, glutathione *S*-transferase. **e**, Phosphorimager quantification of GST–Smt3–Sup35 (open squares) processing to Sup35 (filled squares) in a $[PSI^+]$ strain after labelling with 35 S-methionine, anti-Sup35 immunoprecipitation and SDS–PAGE. Error bars represent standard deviation from two separate experiments. **f**, Sup35 replacement strains exhibit prion-dependent phenotypes. GST, GST–Sup35; GST–Ub, GST–Ub–Sup35; GST–Smt3, GST–Smt3–Sup35.

GFP $^{[psi^-]}$ rather than from the folding of newly translated protein. When Sup35–GFP expression is induced *de novo*, fluorescent cells are not apparent before 40 min (see Supplementary Fig. 7); however zygotes with fluorescent foci were routinely detected within this interval. To monitor the fate of existing Sup35 $^{[psi^-]}$ at earlier times, we observed the mating reaction directly (Fig. 2c, d). Cells about to mate had well-formed mating projections but non-contiguous cytoplasm (Fig. 2c, d; left). Within six minutes, cytoplasmic mixing was evident by the presence of both red and green fluorescence in fused cells (Fig. 2c, d; right). For $[psi^-] \times [psi^-]$ crosses, diffuse green fluorescence persisted, but in $[PSI^+] \times [psi^-]$ matings, mobile green fluorescent foci appeared (Fig. 2c, d, right; see also Supplementary Video 3). Given the rapid repression of P_{MFA1} and the slow maturation of Sup35–GFP (see Supplementary Figs 4 and 7), the appearance of fluorescent foci strongly suggests that mature Sup35–GFP $^{[psi^-]}$ can be converted from one stable form to another.

As Sup35 $^{[PSI^+]}$ is functionally compromised, remodelling of existing Sup35 $^{[psi^-]}$ should induce a translation termination defect. To test this idea, we monitored the appearance of adenine prototrophy in nascent zygotes containing a nonsense mutation in the *ADE1* gene. Although zygotes (see Supplementary Fig. 8) from both $[PSI^+] \times [psi^-]$ and $[psi^-] \times [psi^-]$ crosses continued to divide on complete media, only $[PSI^+] \times [psi^-]$ zygotes divided on media lacking adenine (Fig. 3a). Under these conditions, $[psi^-] \times [psi^-]$ zygotes were unable to form even a single additional bud (Fig. 3a). This immediate phenotypic advantage indicates that Sup35 $^{[psi^-]}$ is rapidly remodelled to the prion form upon fusion of mating partners.

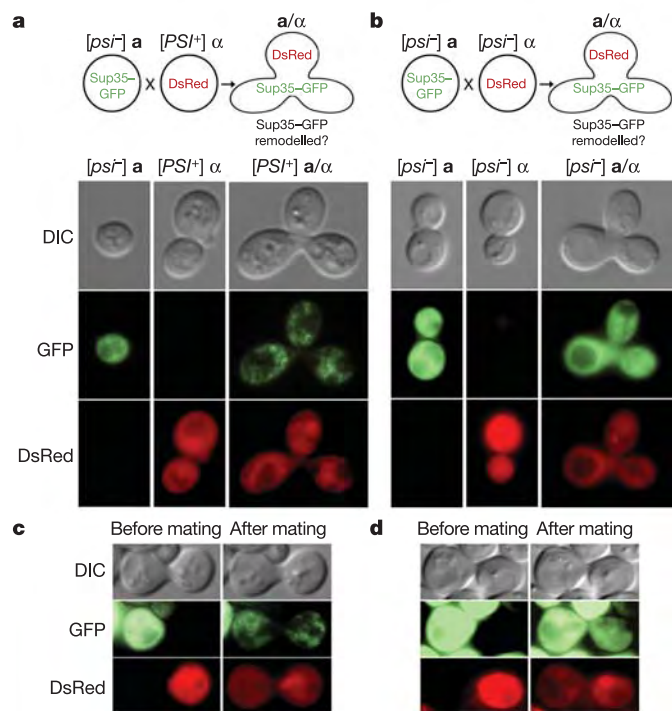


Figure 2 | Existing Sup35 $^{[psi^-]}$ is remodelled after prion transmission by mating. **a**, **b**, $[psi^-]$ *MATa* haploids expressing Sup35–GFP from P_{MFA1} (Sup35–GFP) were mated to either $[PSI^+]$ (**a**) or $[psi^-]$ (**b**) *MATα* haploids constitutively expressing DsRed as indicated in the schematic diagrams. The DIC, GFP and DsRed images are representative of seven independent experiments ($n = 300$). **c**, **d**, Mixtures of mating cells described in **a** and **b**, respectively, were isolated by micromanipulation and visualized by microscopy (before mating, left panels). Within 6 min of the first exposure, the same fields were imaged again (after mating, right panels).

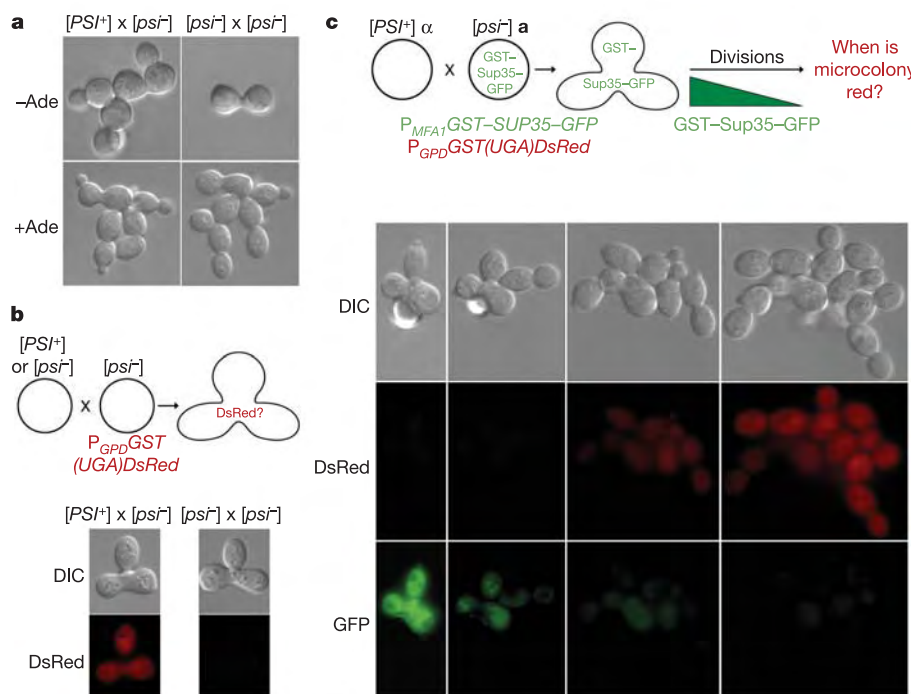


Figure 3 | Phenotypic switching occurs within a single cell cycle after [PSI⁺] transmission by mating. **a**, [PSI⁺] × [psi⁻] or [psi⁻] × [psi⁻] zygotes were isolated by micromanipulation and incubated on media lacking (–Ade) or containing adenine (+Ade). The DIC images are representative of at least five zygotes for each condition. **b**, DIC and DsRed images of zygotes isolated by micromanipulation from the crosses indicated in the schematic diagram. **c**, Zygotes resulting from the crosses indicated in the schematic diagram were isolated by micromanipulation and imaged at 0, 2.5, 6.5, and 9.5 h (left to right). The DIC, DsRed and GFP images are representative of four individual zygotes.

What fraction of the existing Sup35^[psi⁻] must be inactivated to induce this phenotypic switch? To answer this question, we developed a fluorescent reporter GST(UGA)DsRED whose expression is dependent on stop-codon read-through (see Supplementary Fig. 9). When [psi⁻] GST(UGA)DsRED cells are mated to [psi⁻] cells, cloverleaf zygotes (Fig. 3b; see also Supplementary Fig. 10b) remain non-fluorescent; however, in [psi⁻] GST(UGA)DsRED × [PSI⁺] zygotes, red fluorescence is apparent (Fig. 3b; see also Supplementary Fig. 10a). Remodelling of the entire pool of existing Sup35^[psi⁻] is required for this phenotypic switch as red fluorescence is not detected in zygotes when [PSI⁺] cells are mated to a [psi⁻] GST(UGA)DsRED partner that is expressing a functional but non-convertible form of Sup35 (GST–Sup35–GFP) from P_{MFA1} (Fig. 3c). Rather, the [PSI⁺] phenotype gradually emerges in these diploids as the existing GST–Sup35–GFP is diluted by cell division (Fig. 3c). Only after the formation of ~15 buds does the level of read-through in these diploids (as assessed by red fluorescence intensity) reach that observed in zygotes isolated from wild type crosses (Fig. 3c). Therefore, the presence of red fluorescence in wild type [psi⁻] × [PSI⁺] zygotes (Fig. 3b) indicates that nearly all of the active Sup35^[psi⁻] is rapidly remodelled to a functionally compromised state on the introduction of Sup35^[PSI⁺].

As a regulator of the flow of genetic information, Sup35 is poised to profoundly alter cellular phenotype through changes in activity. Such modulation is observed *in vivo*, as Sup35 can adopt multiple, self-perpetuating physical and functional states¹⁸. Intriguingly, propagation of Sup35 forms occurs faithfully enough to direct inheritance of the associated traits, but this process remains sufficiently flexible to allow occasional changes in form. A previously unexplored aspect of this metastability was the temporal appearance of new traits and the molecular mechanism underlying these phenotypic switches. In related work, the prion protein PrP converts to a protease-resistant form post-translationally in mammalian cell lines, but this process is extremely inefficient, leaving open the possibility that this conformer results from an alternate maturation pathway specified earlier in its biogenesis^{19–21}. Our work indicates that the [PSI⁺] prion can arise from the quantitative remodelling of existing Sup35 protein in cells actively propagating the prion form and on

prion transmission through mating. This change in protein state parallels loss of Sup35 activity, suggesting that functionally engaged protein is a substrate for conversion either directly or perhaps at some point during its GTPase cycle²². In either case, the capture of existing protein into prion particles leads to the immediate emergence of new phenotypes. This observation supports previous suggestions that the prion domain of Sup35 has been retained throughout evolution because the epigenetic modulation of Sup35 activity confers selective advantage under appropriate conditions^{18,23}.

The assembly of Sup35^[PSI⁺] complexes both *in vivo* and *in vitro* shares many mechanistic similarities with self-perpetuating changes in the physical states of mammalian proteins associated with severe neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease and Huntington's disease¹⁸. The remodelling of existing proteins from their normal configurations to either a pathogenic intermediate or a potentially protective amyloid²⁴ may have profound effects on the appearance and progression of these diseases. A clear understanding of the *in vivo* protein species competent to undergo physical state transitions, like that described here for the yeast prion [PSI⁺], is an important consideration in the development of therapeutic approaches for these disorders.

METHODS

Strain construction. All strains are derivatives of 74D-694 (*ade1-14 trp1-289 his3Δ200 ura3-52 leu2-3,112*)²⁵. GSTSUP35, GSTSMT3SUP35, GSTUBSUP35, GSTSMT3ΔGGSUP35 and SUP35-GFP strains were constructed by two-step replacement at the SUP35 locus using pRS306-based plasmids digested within P_{SUP35} with *MluI*. Replacements were constructed in wild type [PSI⁺] diploids and verified by polymerase chain reaction (PCR) and segregation after sporulation and tetrad dissection. These strains were converted to [psi⁻] by treatment with 3 mM GdnHCl (ref. 26) unless otherwise indicated. HSP104 was disrupted with *LEU2* by digesting pYSL5 with *PvuI* and *HindIII* before transformation, and the disruption was verified by PCR and loss of [PSI⁺]. [PSI⁺] was cured by overexpression of Hsp104 from P_{GPD} on a 2μ plasmid (p2HG104). *ulp1-333^{ts}* variants were generated by two-step replacement after transformation with *PacI* digested YIP352-*ulp1ts*(3-33) (ref. 13). Replacement was confirmed by inability to grow at 37°C and by the accumulation of Smt3 conjugates by anti-Smt3 immunoblotting (data not shown). Plasmids were integrated after digestion with the indicated endonuclease within the nutritional marker, and integrants were selected by prototrophy: pRS303P_{MFA1} SUP35-GFP (*Eco47III*, *HIS3*);

pRS305P_{LEXAop-GAL1}DsRED (*Clal*, *LEU2*); pRS306P_{GPD}LEXA-GAL4 (*StuI*, *URA3*); pRS303P_{MFA1}GFP-PEST (*Eco47III*, *HIS3*); pRS305P_{MFA1}GST-SUP35-GFP (*PvuMI*, *LEU2*), pRS304P_{GPD}GST(UGA)DsRED (*EcoRV*, *TRP1*).

Plasmid construction. All plasmids were cloned by PCR using standard conditions and verified by sequencing. Primer sequences are available on request. Plasmids for two-step replacement at the *SUP35* locus are based on pRS306. In each, P_{SUP35} is cloned as an *EcoRI*–*BamHI* fragment, and the Sup35 open reading frame (ORF) and fusions are cloned between the *BamHI* and *SacI* sites as either *BglII*–*SacI* fragments (*GST* derivatives) or *BamHI*–*SacI* fragments. GFP is inserted after Sup35 amino-acid 123 in an engineered *SacII* site. This construct differs from a similar recently described Sup35–GFP fusion²⁷ in that our GFP ORF is flanked by linkers encoding three glycine serine repeats on each side. pRS303P_{MFA1} SUP35-GFP contains P_{MFA1} as an *EcoRI*–*BamHI* fragment and SUP35-GFP as a *BamHI*–*SacI* fragment in pRS303. pRS305P_{LEXAop-GAL1}DsRED contains oligomerized LexA binding sites upstream of P_{GAL1} amplified from p8op-lac Z (Clontech) as a *XhoI*–*SacI* fragment and DsRED.T1 (*DsRED*) as a *SacII*–*SacI* fragment in pRS305. p424GPD contains LEXA-GAL4 as a *BamHI*–*EcoRI* fragment amplified from pSH17-4 (S. Hanes, Wadsworth Institute). P_{GPD}LEXA-GAL4 was subcloned as a *KpnI*–*SacI* fragment into pRS306. pRS303P_{MFA1}GFP-PEST contains P_{MFA1} as an *EcoRI*–*BamHI* fragment and GFP-PEST as a *BamHI*–*BglII* fragment subcloned from pSVA13 (ref. 28). p414P_{MET25}MAT α 2 contains MAT α 2 as a *PstI*–*SacI* fragment in p414P_{MET25}. p414P_{MET25}SUP35GFP contains SUP35GFP as a *BamHI*–*NaeI* fragment in p414P_{MET25}. P_{GPD} and the *CYC1* terminator were subcloned as a *SacI*–*KpnI* fragment from p424GPD into pRS304, and GST(UGA)DsRED was subcloned as a *BamHI*–*SacI* fragment to create pRS304P_{GPD}GST(UGA)DsRED. pRS305P_{MFA1}GSTSUP35GFP contains P_{MFA1} as an *EcoRI*–*BamHI* fragment, and GSTSUP35GFP as a *BglII*–*SacI* fragment.

Protein analysis. Lysates were isolated as previously described but with 400 mM NaCl in the lysis buffer²⁹. SDS polyacrylamide gel electrophoresis (SDS–PAGE) and immunoblotting were performed as previously described⁵. Pulse-chase analysis was performed as previously described but labelling was performed in minimal media lacking methionine³⁰. Anti-Sup35 polyclonal rabbit antiserum and protein A-sepharose (Repligen) were used for immunoprecipitation⁵. Quantification was performed on a PhosphorImager.

Yeast mating. Liquid cultures of MAT α and MAT α cells to be mated were grown in minimal media overnight, collected by centrifugation and incubated in media conditioned by overnight cultures of unrelated MAT α and MAT α cells, respectively, for 1 h 20 min, and mixed on solid minimal media containing 2.5 mM adenine to mate at 30 °C. For adenine prototrophy experiments, haploid mating partners were mated as above with the following exceptions: overnight cultures were grown in minimal media containing 15 μ g ml^{−1} adenine; mating reactions were incubated on minimal media containing 5 μ g ml^{−1} adenine and 200 μ g ml^{−1} cytosine; and isolated zygotes were transferred to minimal media lacking adenine for 24 h or containing 2.5 mM adenine for 6.5 h to score growth. Zygotes or haploids conditioned for mating were isolated by micromanipulation. Agar pads with cells were cut from plates and transferred to microscope slides for imaging.

Imaging and digital processing. Zygotes or mating haploids were observed using a Zeiss Axioplan 2 equipped with a $\times 100$ α Plan-FLUAR objective and Hamamatsu-ORCA camera (Hamamatsu Photonics). Images were collected with Openlabs software (Improvision), with a gain of 130, and exposures of 250 ms for GFP, 100 ms for DsRed and 22 ms for DIC. Images of colonies were collected with Kodak DC290 digital camera and Kodak 1D 3.5.4 software (Scientific Imaging Systems). Images were processed, and fluorescent images were pseudocoloured in Adobe Photoshop 7.0. Supplementary movies were collected and processed in Openlabs software.

Received 4 March; accepted 24 June 2005.

- Weissmann, C. The state of the prion. *Nature Rev. Microbiol.* **2**, 861–871 (2004).
- Uptain, S. M. & Lindquist, S. Prions as protein-based genetic elements. *Annu. Rev. Microbiol.* **56**, 703–741 (2002).
- Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144 (1982).
- Serio, T. R. & Lindquist, S. L. [*PSI*⁺]: an epigenetic modulator of translation termination efficiency. *Annu. Rev. Cell Dev. Biol.* **15**, 661–703 (1999).
- Patino, M. M., Liu, J. J., Glover, J. R. & Lindquist, S. Support for the prion hypothesis for inheritance of a phenotypic trait in yeast. *Science* **273**, 622–626 (1996).
- Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. Propagation of the yeast prion-like [*PSI*⁺] determinant is mediated by

oligomerization of the SUP35-encoded polypeptide chain release factor. *EMBO J.* **15**, 3127–3134 (1996).

- Cox, B. [*PSI*], a cytoplasmic suppressor of super-suppression in yeast. *Heredity* **20**, 505–521 (1965).
- Dagkesamanskaya, A. R., Kushnirov, V. V., Paushkin, S. V. & Ter-Avanesyan, M. D. Fusion of glutathione S-transferase with the N-terminus of yeast Sup35 protein inhibits its prion-like properties. *Genetika* **33**, 610–615 (1997).
- Glover, J. R. et al. Self-seeded fibers formed by Sup35, the protein determinant of [*PSI*⁺], a heritable prion-like factor of *S. cerevisiae*. *Cell* **89**, 811–819 (1997).
- Paushkin, S. V., Kushnirov, V. V., Smirnov, V. N. & Ter-Avanesyan, M. D. *In vitro* propagation of the prion-like state of yeast Sup35 protein. *Science* **277**, 381–383 (1997).
- King, C. Y. et al. Prion-inducing domain 2–114 of yeast Sup35 protein transforms *in vitro* into amyloid-like filaments. *Proc. Natl Acad. Sci. USA* **94**, 6618–6622 (1997).
- Griffith, J. Self-replication and Scrapie. *Nature* **215**, 1043–1044 (1967).
- Li, S. J. & Hochstrasser, M. A new protease required for cell-cycle progression in yeast. *Nature* **398**, 246–251 (1999).
- Amerik, A. Y. & Hochstrasser, M. Mechanism and function of deubiquitinating enzymes. *Biochim. Biophys. Acta* **1695**, 189–207 (2004).
- Finley, D., Bartel, B. & Varshavsky, A. The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* **338**, 394–401 (1989).
- Johnson, E. S., Schwienshorst, I., Dohmen, R. J. & Blobel, G. The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aoi1p/Uba2 heterodimer. *EMBO J.* **16**, 5509–5519 (1997).
- Galgoczy, D. J. et al. Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **101**, 18069–18074 (2004).
- Chien, P., Weissman, J. S. & DePace, A. H. Emerging principles of conformation-based prion inheritance. *Annu. Rev. Biochem.* **73**, 617–656 (2004).
- Borchelt, D. R., Scott, M., Taraboulos, A., Stahl, N. & Prusiner, S. B. Scrapie and cellular prion proteins differ in their kinetics of synthesis and topology in cultured cells. *J. Cell Biol.* **110**, 743–752 (1990).
- Borchelt, D. R., Taraboulos, A. & Prusiner, S. B. Evidence for synthesis of scrapie prion proteins in the endocytic pathway. *J. Biol. Chem.* **267**, 16188–16199 (1992).
- Caughey, B. & Raymond, G. J. The scrapie-associated form of PrP is made from a cell surface precursor that is both protease- and phospholipase-sensitive. *J. Biol. Chem.* **266**, 18217–18223 (1991).
- Frolova, L. et al. Eukaryotic polypeptide chain release factor eRF3 is an eRF1- and ribosome-dependent guanosine triphosphatase. *RNA* **2**, 334–341 (1996).
- True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
- Caughey, B. & Lansbury, P. T. Protofibrils, pores, fibrils and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26**, 267–298 (2003).
- Chernoff, Y. O., Lindquist, S. L., Ono, B., Inge-Vechtomov, S. G. & Lieberman, S. W. Role of the chaperone protein Hsp104 in propagation of the yeast prion-like factor [*PSI*⁺]. *Science* **268**, 880–884 (1995).
- Tuite, M. F., Mundy, C. R. & Cox, B. S. Agents that cause a high frequency of genetic change from [*PSI*⁺] to [*psi*[−]] in *Saccharomyces cerevisiae*. *Genetics* **98**, 691–711 (1981).
- Song, Y. et al. Role for Hsp70 Chaperone in *Saccharomyces cerevisiae* Prion Seed Replication. *Eukaryot. Cell* **4**, 289–297 (2005).
- Mateus, C. & Avery, S. V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* **16**, 1313–1323 (2000).
- Ness, F., Ferreira, P., Cox, B. S. & Tuite, M. F. Guanidine hydrochloride inhibits the generation of prion “seeds” but not prion protein aggregation in yeast. *Mol. Cell. Biol.* **22**, 5593–5605 (2002).
- Chen, P., Johnson, P., Sommer, T., Jentsch, S. & Hochstrasser, M. Multiple ubiquitin-conjugating enzymes participate in the *in vivo* degradation of the yeast MAT α 2 repressor. *Cell* **74**, 357–369 (1993).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work is dedicated to the memory of A. Serio, Jr. We thank S. Lindquist, M. Hochstrasser, J. Laney, B. Glick and S. Hanes for reagents and discussions, J. Laney, K. Mowry, A. Jacobson, J. Sedivy, A. Landy, M. McKeown, A. DeLong, M. Strbuncelj, S. Langseth and Z. Zinn for comments on the manuscript, and R. Lesiak and J. Nathanson for graphics assistance. This research was supported by the National Cancer Institute and the Pew Scholars Program in the Biomedical Sciences.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.R.S. (Tricia_Serio@Brown.edu).

LETTERS

Amyloid-like fibrils of ribonuclease A with three-dimensional domain-swapped and native-like structure

Shilpa Sambashivan¹, Yanshun Liu^{1†}, Michael R. Sawaya¹, Mari Gingery¹ & David Eisenberg¹

Amyloid or amyloid-like fibrils are elongated, insoluble protein aggregates, formed *in vivo*¹ in association with neurodegenerative diseases or *in vitro*² from soluble native proteins, respectively. The underlying structure of the fibrillar or 'cross- β ' state has presented long-standing, fundamental puzzles of protein structure. These include whether fibril-forming proteins have two structurally distinct stable states, native and fibrillar, and whether all or only part of the native protein refolds as it converts to the fibrillar state. Here we show that a designed amyloid-like fibril of the well-characterized enzyme RNase A contains native-like molecules capable of enzymatic activity. In addition, these functional molecular units are formed from a core RNase A domain and a swapped complementary domain. These findings are consistent with the zipper-spine model³ in which a cross- β spine is decorated with three-dimensional domain-swapped functional units, retaining native-like structure.

The discovery that numerous native proteins can be converted to the amyloid-like fibrillar state⁴ seems to challenge a central pillar of protein science. This is Anfinsen's 'thermodynamic hypothesis'⁵, which states that protein sequence determines a unique structure of lowest free energy. Proteins in the fibrillar, amyloid-like state have common properties⁶: these include an elongated, fibrillar morphology; a 'cross- β ' diffraction pattern, indicative of β -strands perpendicular to the fibril axis and β -sheets parallel to the axis; and binding of flat dyes, giving common tinctural properties. From this it has been inferred² that "amyloid is a generic structural form of

proteins", stabilized by peptide backbone bonding. Is this proposed generic amyloid, in which backbone bonding dominates a second type of structure, in conflict with the thermodynamic hypothesis in which side-chain interactions determine a single stable state for each protein?

The fibrillar nature of the amyloid state has complicated structural studies. Cryoelectron microscopy has been able to resolve helical protofilaments forming the fibril^{7,8}, and in one case it has been suggested that native domains must unfold in the amyloid form⁷. In some studies (for example, ref. 9), domains seem to decorate the surface of the spine. A currently popular molecular model for the amyloid-like protofibrils depicts part or all of the native protein as refolding into a parallel, left-handed β -helix^{10–13}. Another type of model invokes three-dimensional (3D) domain swapping to account for the protein self-specificity of amyloid interactions^{14,15} and cohesion. For cystatin amyloid fibrils, domain swapping has been supported by biochemical experiments^{16–19}. In short, despite progress, there is as yet little consensus on major questions of amyloid structure, including whether native-like domains are retained or are refolded in the fibrils.

Here we address this question with RNase A, the same protein whose spontaneous refolding in solution led to the original thermodynamic hypothesis²⁰. The RNase A molecule has three properties that make it convenient for answering this question. First, it is tightly cross-linked by four disulphide bonds, severely restricting conformational change (Fig. 1a). Second, RNase A forms two types of 3D

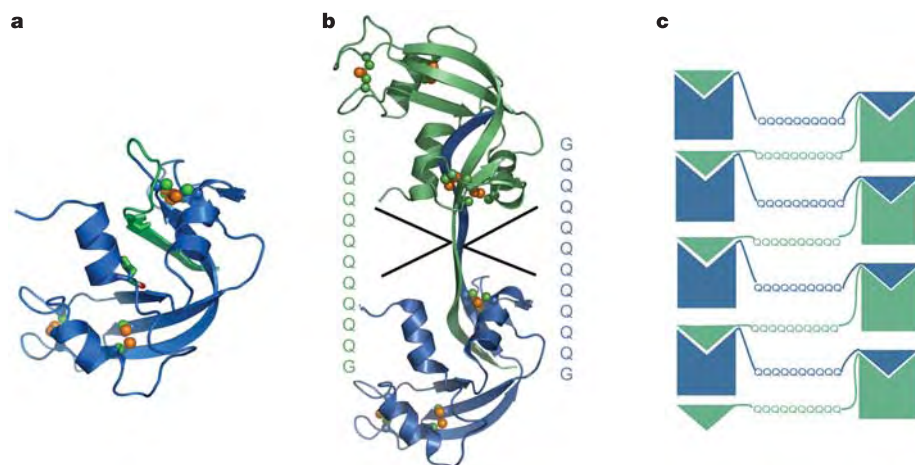


Figure 1 | RNase A monomer and C-terminal domain-swapped dimer and the 3D domain-swapped zipper-spine model. **a**, The RNase A monomer is stabilized by four disulphide bonds Cys 26–Cys 84, Cys 40–Cys 95, Cys 58–Cys 110 and Cys 65–Cys 72, hindering conformational changes. His 12 in the core of the protein and His 119 on the β -strand that is swapped (shown by sticks) are active-site residues that we mutate to test for activity by complementation. **b**, The C-terminal domain-swapped dimer is formed by exchanging the C-terminal β -strands between two monomers. The hinge loop (residues 112–115) has been expanded by inserting the sequence -GQ₁₀G-. **c**, Diagram of amyloid-like fibril formation in RNase A with Q₁₀ expansion, leading to a runaway domain swap. The Q₁₀-H12A mutants are shown in blue and the Q₁₀-H119A mutants in green. Domain swapping between two mutants complements active sites.

¹Howard Hughes Medical Institute, UCLA-DOE Institute for Genomics and Proteomics, Box 951570, UCLA, Los Angeles, California 90095-1570, USA. [†]Present address: Division of Chemistry and Chemical Engineering, Caltech, Pasadena, California 91125, USA.

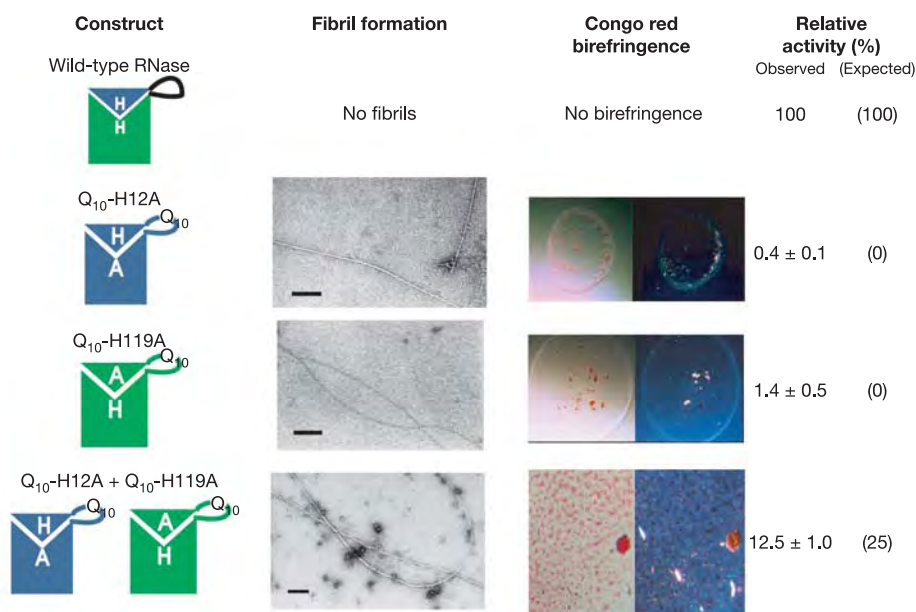


Figure 2 | Properties of the RNase A amyloid-like fibrils. The hinge-loop region of wild-type RNase A that connects the C-terminal β -strand (triangle in the diagrams in column 1) to the protein core is expanded with the -GQ₁₀G- motif to generate amyloid-forming RNase A mutants. Two inactive RNase A mutants are formed by replacing His 12 or His 119 with Ala. Wild-type RNase A does not form fibrils and has a fully functional active site (row 1, columns 2 and 4). The Q₁₀-H12A, Q₁₀-H119A and Q₁₀-H12A + Q₁₀-

H119A constructs all form amyloid-like fibrils (column 2) and bind Congo red with the characteristic apple-green birefringence (column 3). The Q₁₀-H12A + Q₁₀-H119A fibrils (row 4, column 4) have significantly higher activity than fibrils of Q₁₀-H12A (row 2, column 4) and Q₁₀-H119A (row 3, column 4) alone. This is a result of complementation of active sites by domain swapping. The expected activity of each of the constructs is given in parentheses. Scale bar (column 2), 200 nm.

domain swaps when freeze-dried from 40% acetic acid^{3,21,22}. In one of these (Fig. 1b) it exchanges its carboxy-terminal β -strand with that of an identical molecule. From the structure of the C-terminal swapped RNase A dimer, it has been speculated³ that if the hinge loop (residues 112–115) connecting the core domain (residues 1–111) with the swapped domain (residues 116–124) were expanded by insertion of an amyloidogenic segment, then a domain-swapped amyloid-like fibril might form, as depicted schematically in Fig. 1c. Third, one of the two catalytic His residues of RNase A (at position 12) is on the

core domain of the model (Fig. 1c), whereas the other (at position 119) is on the swapped domain. This segregation of the catalytic His residues on different domains offers the possibility of forming an active, complementary dimer from two inactive RNase A monomers, as explained below.

Before the determination of the structure of RNase A, it was found²¹ that when inactive RNase A, alkylated at His 12, was freeze-dried with inactive RNase A, alkylated at His 119, the resulting oligomers had restored activity. The conclusion was that “the

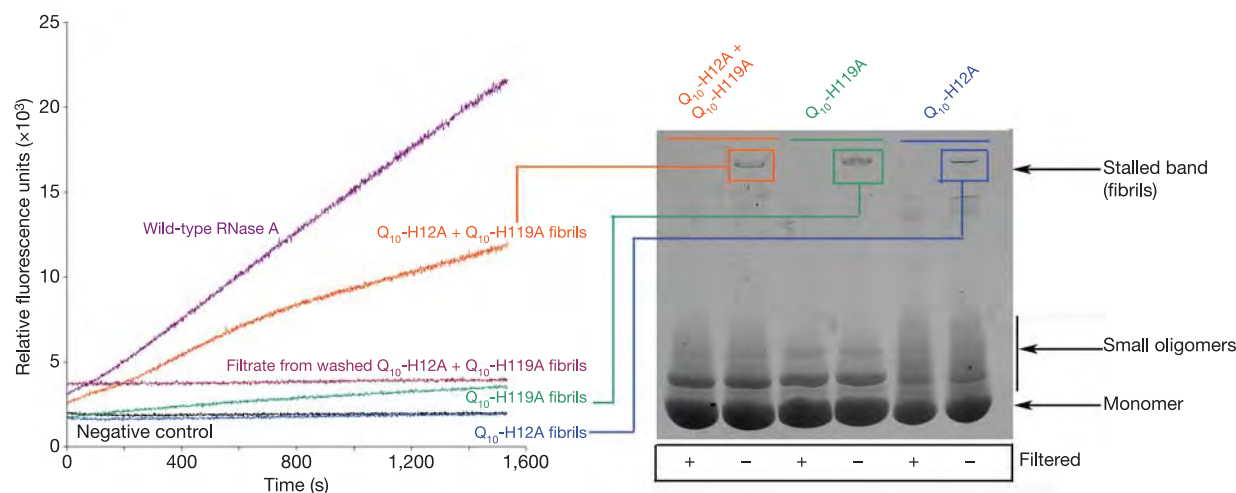


Figure 3 | Fluorescence-based activity assay to examine whether the RNase A amyloid-like fibrils are domain-swapped. The amyloid-like fibrils are separated from smaller oligomers by subjecting the samples to electrophoresis on a non-denaturing gel (right). The fibril-containing stalled bands (confirmed by in-gel Congo red assays) are assayed for activity in solution by a fluorescence assay (left). The slope of each of the traces corresponds to the amount of product formed per unit time. Thus the slope

per mg of the enzyme corresponds to the specific activity of the enzyme. The Q₁₀-H12A (blue) and Q₁₀-H119A fibrils (green) have slopes comparable to the negative control (black) and thus a specific activity comparable to background. The Q₁₀-H12A + Q₁₀-H119A sample (orange) has a much steeper slope and hence a significantly higher specific activity (Fig. 2, column 4) than background.

histidine residues at positions 119 and 12 both form part of the active site of the enzyme", and foreshadowed the discovery of 3D domain swapping by 30 years. Here we use a variation of this experiment, with either His 12 or His 119 replaced by Ala in hinge-loop-expanded RNase molecules, to demonstrate native-like domains and domain swapping in amyloid-like RNase A fibrils.

We converted RNase A to the fibrillar form by expanding the hinge loop after Gly 112 with an insert of ten glutamine residues flanked by a glycine residue on each side (GQ₁₀G). When freeze-dried from 50% acetic acid, the constructs with the Q₁₀ expansions form fibrils (Fig. 2, column 2) that bind the dye Congo red and show 'apple green' birefringence characteristic of amyloid (Fig. 2, column 3). One of these amyloid-like fibrils (Q₁₀-H119A; Fig. 2, row 3) was subjected to X-ray fibre diffraction and showed the characteristic strong reflection at 4.8 Å and the diffuse reflection at 11 Å resolution (Supplementary Fig. S1). These observations indicate that the RNase A molecules with Q₁₀ expansions in the C-terminal hinge loop form amyloid-like fibrils on freeze-drying. RNase A with a GQ₇G or a GNNQQNY (an amyloidogenic motif from the yeast Sup35 prion) expansion also forms amyloid-like fibrils. In contrast, wild-type RNase A did not form fibrils in more than a dozen experiments, as judged by both electron microscopy and silver-stained non-denaturing gels (Supplementary Fig. S2 and Supplementary Table 1); nor did a Gly₉ expansion of RNase A, having nine glycine residues after Gly 112, form fibrils.

The crystal structure at 1.8 Å resolution of the Q₁₀-H119A RNase A before freeze-drying reveals an RNase A molecule virtually identical to the wild type, except that the hinge loop, including the Q₁₀ expansion, is disordered and invisible in the electron density (data not shown). This shows that the Q₁₀ expansion does not alter the basic features of monomeric RNase A, as expected from the four disulphide bridges that stabilize the monomer.

What happens to the structure of Q₁₀-expanded RNase A on freeze-drying, as fibrils form? One possibility is that the Q₁₀ segments stack together to form a β -sheet, with the RNase A molecules otherwise unaltered. A second possibility is envisaged in the zipper-spine model (Fig. 1c)³: the C-terminal β -strand of each RNase A monomer breaks its non-covalent bonds with the core domain, exposing the Q₁₀ expansion more fully, and the C-terminal β -strand of another molecule swaps in to take its place, forming a domain-swapped functional unit of RNase A. The Q₁₀ expansion loops can then stack into a zipper spine^{3,23}, with the core domains and C-terminal β -strands complementing each other at the periphery of the spine.

We distinguished between these two possibilities by testing the activities of fibrils formed from inactive Q₁₀-H12A RNase A and Q₁₀-H119A RNase A, alone or mixed. The fibrils formed from each of the inactive RNase A molecules are themselves inactive (Fig. 2, rows 2 and 3). However, when these two inactive RNase A constructs are mixed and then freeze-dried, the resulting fibrils now show significant RNase enzymatic activity (Fig. 2, bottom row). This activity must be the result of complementation between RNase A molecules, composite active sites having been formed by domain swapping between the H12A and H119A mutants (as shown schematically in Fig. 1c). That is, the RNase A amyloid-like fibrils are apparently made up of 3D domain-swapped RNase A molecules.

There are two conceivable alternative explanations for the activity attributed to the amyloid-like fibrils, both of which we have ruled out. One alternative is that domain-swapped complementary small oligomers cling to the fibrils in a non-specific manner, resulting in apparently active fibrils. Alternatively, there might be large, non-fibrillar, domain-swapped, active aggregates in the stalled band (Fig. 3, right panel). That is, the stalled band might contain two species: inactive amyloid-like fibrils and active domain-swapped aggregates. From a series of control experiments (see Supplementary Information) we conclude that that activity of the mixed fibrils (Fig. 2, bottom row, and Fig. 3, orange trace) comes from complemented

active sites in the fibrils and not from smaller clinging oligomers or larger domain-swapped non-fibrillar aggregates. This means that the RNase A amyloid-like fibrils contain native-like, domain-swapped functional units.

We propose a model (Fig. 4) for the RNase A amyloid-like fibril, based on our biochemical data, suggesting domain-swapped functional units. In this model the spine of the structure is a twisted pair of antiparallel β -sheets. Each β -strand is the 10-glutamine expansion from the hinge-loop segment of RNase A. The β -strands are stacked 4.88 Å apart along the fibril axis (Fig. 4c), as in the crystal structure of the cross- β spine²³. A 7° twist is introduced between successive Q₁₀ segments of the spine to account for the twisted appearance of the fibrils (Fig. 2). Each Q₁₀ expansion forms hydrogen bonds to identical segments both above and below within the sheet but not between the sheets. Instead, the Q₁₀ side chains from one sheet interdigitate with those from the other sheet, forming what we term a steric zipper²³ (Fig. 4b). The model depicts the native fold of RNase A as being essentially retained with only a small segment of the protein (the Q₁₀-hinge loop) forming the cross β -spine.

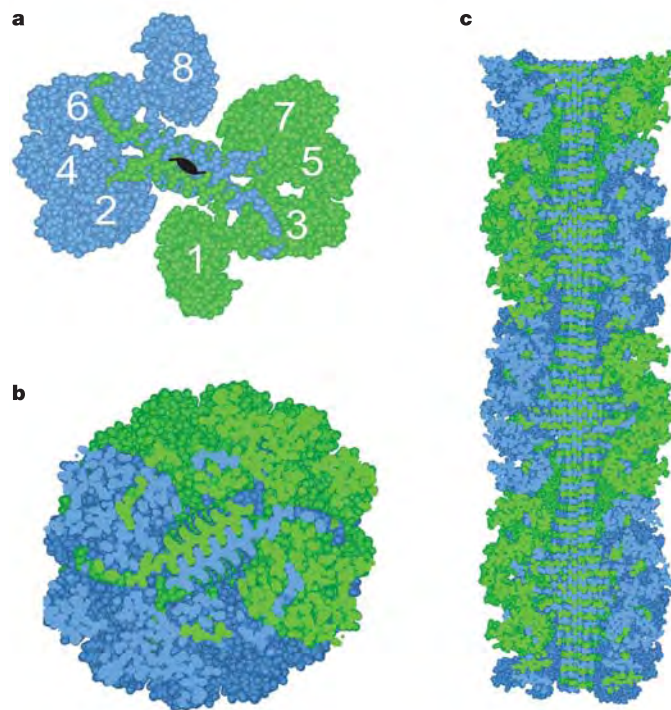


Figure 4 | Domain-swapped zipper-spine model for the RNase A protofibril. **a**, The model is a 'runaway' domain swap between the RNase A monomers, with swaps occurring within one half protofibril but not between half protofibrils. Monomers 1–4 compose half the protofibrillar unit and are coloured as in Fig. 1c to emphasize domain swapping. The C-terminal β -strand of monomer 1 swaps into monomer 2, monomer 2 swaps into monomer 3, and monomer 3 swaps into 4, rising along the axis of the fibril. Q₁₀ segments from these monomers form one antiparallel β -sheet in the spine. Monomers 5–8 form the other β -sheet, related to monomers 1–4 by a 2₁ axis along the fibril. Eight RNase A monomers comprise the asymmetric unit of the fibril. A similar model can be built from domain-swapped dimers; the currently available data do not favour one of these models over the other. **b**, The protofibril cross-section reveals the steric zipper, the interdigitation of Gln side chains in the spine of the fibril, modelled on the structure of GNNQQNY²³. **c**, A cut-away view perpendicular to the fibril axis reveals the stacking of hydrogen-bonded Q₁₀ β -strands (4.88 Å apart) in the spine. The spine is largely shielded from solvent by the tight packing of globular domains around the periphery. The fibril model ranges from 100 to 140 Å in diameter, which agrees with the diameter of the fibrils obtained from electron microscopy images (Fig. 2, column 2).

The model of RNase A amyloid of Fig. 4 is compatible with the thermodynamic hypothesis⁵, provided that we take a wider than usual view of what is meant by “the three-dimensional structure of a native protein in its physiological milieu”⁵. In domain-swapped structures, such as the RNase A dimer of Fig. 1b or the model of Fig. 4, RNase A molecules differ considerably in conformation from the monomer (compare the chains in Fig. 1a and Fig. 1b). This is so because the conformations of the hinge loops differ, leading to greatly differing relationships of the core and swapped domains in the two structures. But if we focus on the ‘functional unit’²⁴ consisting of the pair of complementary domains from two RNase A molecules, the structure of the functional unit of the dimer is essentially the same as the structure of the monomer. In fact, the functional unit supports enzymatic activity, taken as the hallmark of the native state in the original experiments²⁰. Thus, the functional unit of the domain-swapped molecule can be considered ‘native-like’ if not actually ‘native’. In other words, to the extent that the monomeric and domain-swapped functional units have the same structure, this common structure is determined by the sequence, and the hypothesis holds. We note that others have given evidence that cystatins^{16–19} and β_2 -microglobulin¹⁴ are domain-swapped, so that the amyloid-like fibrils of these proteins might also be compatible with the thermodynamic hypothesis. Still another amyloid-like fibril, formed from human lysozyme²⁵, was found to lack enzymatic activity, but the lack of activity would not rule out domain swapping if the hinge loop, which is changed in conformation on swapping, includes one or more active site residues.

Thus, we have designed an amyloid-like fibril of the stable enzyme RNase A with a Q₁₀ expansion between its core domain and its C-terminal β -strand. The fibril has RNase enzymatic activity, showing that it contains native-like domains. By mixing two inactive mutant forms of Q₁₀-RNase A we find that activity is restored, showing that domains are swapped in the fibril. 3D domain swapping can account for the protein self-specificity observed in amyloid fibrils and for at least part of the cohesion of the fibril, and part of the kinetic barrier between monomer and fibril²⁶. A speculative domain-swapped helical model is proposed for the RNase A fibril, which is compatible with Anfinsen’s thermodynamic hypothesis⁵ that protein sequence determines the most stable structure. For the RNase A amyloid-like fibril there is no need to invoke a distinctly different, generic structural form of protein, other than for a small segment that forms the β -sheet spine.

METHODS

Strain construction. A wild-type RNase A clone from R. T. Raines was subcloned into the *SalI*–*NcoI* sites of the pET-32b vector system. The H12A and H119A single mutants were generated by site-directed mutagenesis. The GQ₁₀G, GQ₇G and G₉ sequences were inserted into the hinge-loop region (after Gly 112) of the mutant RNase A by two-step insertion polymerase chain reaction. The GNNQQNY expansion was formed by inserting NQQNYGG after Asn 113 (GN corresponds to residues 112 and 113 of the wild-type protein and NQQNYGG corresponds to the inserted sequence). See Supplementary Information for details of expression and purification.

Fibril formation. The purified mutants (pH 7.0) were incubated in 50% acetic acid at concentrations from 0.15 to 1.5 mM, followed by freezing the samples in dry ice and freeze-drying overnight. The freeze-dried samples were resuspended in 30–40 μ l of water and stored on the benchtop.

Activity assays. Fibril bands that stalled in the native stacking gel (see the text) were excised and soaked overnight in water at pH 7.0. These solutions were assayed in duplicate for RNase A activity with the fluorescence kit from Ambion. The negative control was the substrate in the assay buffer. The positive control was wild-type RNase A. See Supplementary Information for details.

Model building. The model-building procedures are described in Supplementary Information.

Received 18 January; accepted 16 June 2005.

1. Koo, E. H., Lansbury, P. T. Jr & Kelly, J. W. Amyloid diseases: abnormal protein aggregation in neurodegeneration. *Proc. Natl Acad. Sci. USA* **96**, 9989–9990 (1999).
2. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332 (1999).
3. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. A domain-swapped RNase A dimer with implications for amyloid formation. *Nature Struct. Biol.* **8**, 211–214 (2001).
4. Sunde, M. *et al.* Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **273**, 729–739 (1997).
5. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
6. Cohen, A. A. & Calkins, E. Electron microscopic observations on a fibrous component in amyloid of diverse origins. *Nature* **183**, 1202–1203 (1959).
7. Jimenez, J. L. *et al.* Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.* **18**, 815–821 (1999).
8. Jimenez, J. L. *et al.* The protofilament structure of insulin amyloid fibrils. *Proc. Natl Acad. Sci. USA* **99**, 9196–9201 (2002).
9. Kishimoto, A. *et al.* β -Helix is a likely core structure of yeast prion Sup amyloid fibrils. *Biochem. Biophys. Res. Commun.* **315**, 739–745 (2004).
10. Govaerts, C., Wille, H., Prusiner, S. B. & Cohen, F. E. Evidence for assembly of prions with left-handed β -helices into trimers. *Proc. Natl Acad. Sci. USA* **101**, 8342–8347 (2004).
11. Perutz, M. F., Finch, J. T., Berriman, J. & Lesk, A. Amyloid fibrils are water-filled nanotubes. *Proc. Natl Acad. Sci. USA* **99**, 5591–5595 (2002).
12. Pickersgill, R. W. A primordial structure underlying amyloid. *Structure* **11**, 137–138 (2003).
13. Wetzel, R. Ideas of order for amyloid fibril structure. *Structure* **10**, 1031–1036 (2002).
14. Eakin, C. M., Attenello, F. J., Morgan, C. J. & Miranker, A. D. Oligomeric assembly of native-like precursors precedes amyloid formation by β -2 microglobulin. *Biochemistry* **43**, 7808–7815 (2004).
15. Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv. Protein Chem.* **50**, 61–122 (1997).
16. Janowski, R. *et al.* Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. *Nature Struct. Biol.* **8**, 316–320 (2001).
17. Nilsson, M. *et al.* Prevention of domain swapping inhibits dimerization and amyloid fibril formation of cystatin C: use of engineered disulfide bridges, antibodies, and carboxymethylpapain to stabilize the monomeric form of cystatin C. *J. Biol. Chem.* **279**, 24236–24245 (2004).
18. Sanders, A. *et al.* Cystatin forms a tetramer through structural rearrangement of domain-swapped dimers prior to amyloidogenesis. *J. Mol. Biol.* **336**, 165–178 (2004).
19. Staniforth, R. A. *et al.* Three-dimensional domain swapping in the folded and molten-globule states of cystatins, an amyloid-forming structural superfamily. *EMBO J.* **20**, 4774–4781 (2001).
20. Sela, M., White, F. H. & Anfinsen, C. B. Reductive cleavage of disulfide bridges in ribonuclease. *Science* **125**, 691–692 (1957).
21. Crestfield, A. M., Stein, W. H. & Moore, S. Properties and conformation of the histidine residues at the active site of ribonuclease. *J. Biol. Chem.* **228**, 2421–2428 (1963).
22. Gotte, G., Bertoldi, M. & Libonati, M. Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme. *Eur. J. Biochem.* **265**, 680–687 (1999).
23. Nelson, R. *et al.* Structure of the amyloid spine. *Nature* **435**, 773–777 (2005).
24. Liu, Y. & Eisenberg, D. 3D domain swapping: as domains continue to swap. *Protein Sci.* **11**, 1285–1299 (2002).
25. Booth, D. R. *et al.* Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787–793 (1997).
26. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455–2468 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Phillips for electron microscopy, M. Bennett, T. Yeates and D. H. Anderson for discussion, and M. Nebahacova and L. Simpson for assistance with blue-native gels. This study received financial support from the NIH, NSF and the Howard Hughes Medical Institute.

Author Information The structure of Q₁₀-H119A RNase A has been deposited in the Protein Data Bank with accession code 2APQ. The model of Fig. 4 has been deposited in the Protein Data Bank with accession code 2APU. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to D.E. (david@mbi.ucla.edu).

LETTERS

Mutations in *NOTCH1* cause aortic valve disease

Vidu Garg^{1,5}, Alecia N. Muth^{1†}, Joshua F. Ransom^{1†}, Marie K. Schluterman¹, Robert Barnes^{3,4}, Isabelle N. King^{1,5†}, Paul D. Grossfeld⁶ & Deepak Srivastava^{1,2,4,5†}

Calcification of the aortic valve is the third leading cause of heart disease in adults¹. The incidence increases with age, and it is often associated with a bicuspid aortic valve present in 1–2% of the population². Despite the frequency, neither the mechanisms of valve calcification nor the developmental origin of a two, rather than three, leaflet aortic valve is known. Here, we show that mutations in the signalling and transcriptional regulator *NOTCH1* cause a spectrum of developmental aortic valve anomalies and severe valve calcification in non-syndromic autosomal-dominant human pedigrees. Consistent with the valve calcification phenotype, *Notch1* transcripts were most abundant in the developing aortic valve of mice, and *Notch1* repressed the activity of *Runx2*, a central transcriptional regulator of osteoblast cell fate. The hairy-related family of transcriptional repressors (Hrt), which are activated by *Notch1* signalling, physically

interacted with *Runx2* and repressed *Runx2* transcriptional activity independent of histone deacetylase activity. These results suggest that *NOTCH1* mutations cause an early developmental defect in the aortic valve and a later de-repression of calcium deposition that causes progressive aortic valve disease.

Abundant evidence suggests a major inherited component to the aetiology of aortic valve disease in children and adults^{3,4}. The most severe type of aortic valve obstruction in children results in failure of the fetal left ventricle to grow, a condition known as hypoplastic left heart syndrome. About 10% of relatives of hypoplastic left heart syndrome patients have bicuspid aortic valve, often undiagnosed, suggesting a common genetic aetiology with phenotypic heterogeneity⁵. The valve calcification often observed in bicuspid aortic valve is a result of inappropriate activation of osteoblast-specific gene expression⁶, but the mechanism is unknown.

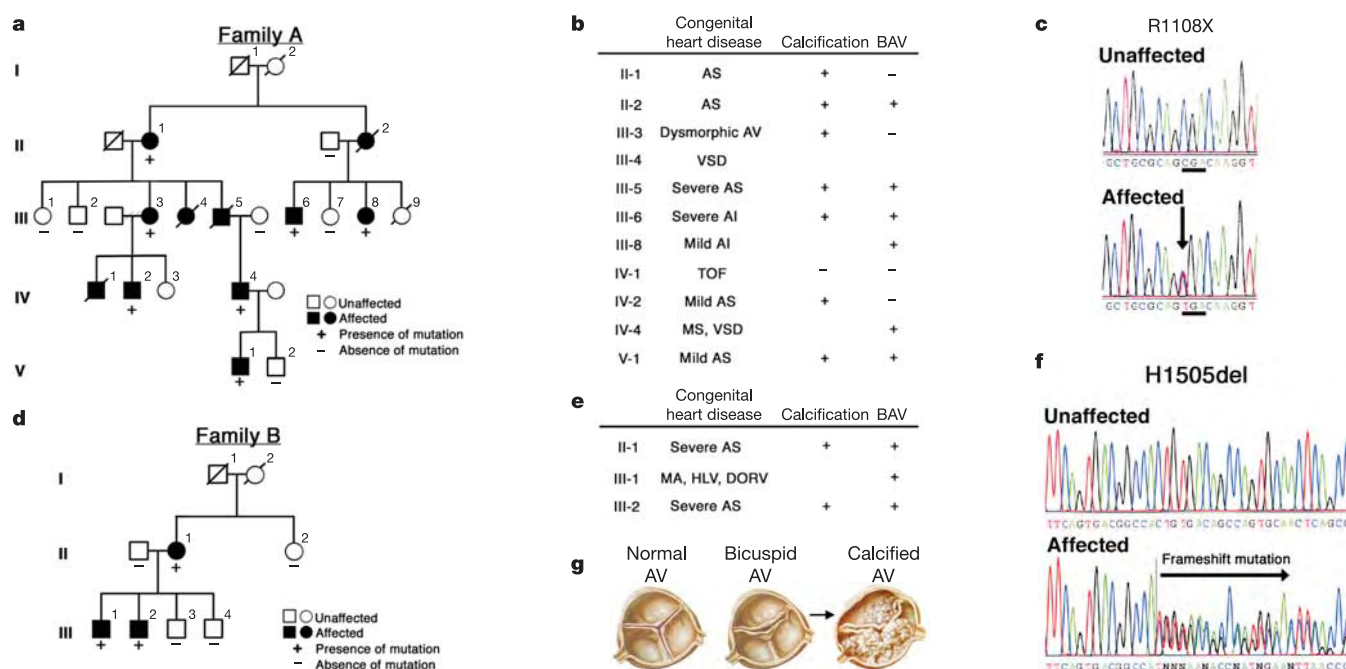


Figure 1 | *NOTCH1* mutations segregate with familial aortic valve disease. **a**, Kindred with five generations (indicated with Roman numerals) affected by congenital heart disease and valve calcification. Participating members of each generation are indicated numerically. Deceased family members (slash) were unavailable for mutation analysis. Squares, males; circles, females. **b**, Cardiac phenotype in affected family members. AI, aortic insufficiency; AS, aortic stenosis; AV, aortic valve; BAV, bicuspid aortic valve;

TOF, tetralogy of Fallot; VSD, ventricular septal defect. **c**, Sequence chromatogram of affected family members. **d**, Kindred with three members affected by congenital heart disease. **e**, Cardiac phenotype of family B. DORV, double-outlet right ventricle; HL, hypoplastic left ventricle; MA, mitral atresia; MS, mitral stenosis. **f**, Sequence chromatogram of affected members in family B. **g**, Schematic of normal trileaflet aortic valve, bicuspid aortic valve and calcified aortic valve.

¹Departments of Pediatrics, ²Molecular Biology and ³Internal Medicine, and ⁴the McDermott Center for Human Growth and Development, 6000 Harry Hines Boulevard, Room NA8.124, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9148, USA. ⁵Children's Medical Center, Dallas, Texas 75235, USA. ⁶Department of Pediatrics, Division of Cardiology, University of California, San Diego 92123, USA. †Present address: Gladstone Institute of Cardiovascular Disease and Department of Pediatrics, University of California, San Francisco, 1650 Owens Street, San Francisco, California 94158, USA.

We identified a family of European–American descent spanning five generations with 11 cases of congenital heart disease (Fig. 1a). Clinical evaluations demonstrated autosomal-dominant inheritance of congenital heart disease. Nine affected family members had aortic valve disease (Fig. 1b,g). In eight, an abnormal aortic valve was the only cardiac malformation; six had bicuspid aortic valve, and seven developed calcific aortic stenosis, including three cases in the setting of a three leaflet valve. One family member (IV-4) had an associated abnormal mitral valve, resulting in mitral stenosis, and a ventricular septal defect. An isolated ventricular septal defect or tetralogy of Fallot with a bicuspid pulmonary valve was identified in two other affected family members (III-4 and IV-1, respectively). Four family members have required aortic valve replacement for severe calcification. No cardiac conduction abnormalities, neurological deficits, or other birth defects were identified. Detailed clinical phenotype information for this family is shown in Supplementary Fig. 1a.

A genome-wide scan of available family members revealed linkage of the congenital heart disease phenotype to a single locus on chromosome 9q34–35 between D9S1826 and D9qter (logarithm of odds (LOD) score, 3.5, $\theta = 0$), spanning approximately 3 megabases (~ 9 cM) (for haplotype data, see Supplementary Fig. 2). Review of 30 known (and 57 predicted) genes revealed *NOTCH1*, which encodes a transmembrane receptor (2,556 amino acids) that functions in a highly conserved intracellular signalling pathway involved in cellular differentiation, cell fate and lateral inhibition⁷. Direct sequencing of *NOTCH1* in an affected patient revealed a heterozygous C-to-T transition of nucleotide 3322 that predicted a premature stop codon instead of arginine at position 1108 in the extracellular domain (Fig. 1c). All affected subjects who were clinically evaluated had the R1108X mutation, suggesting autosomal-dominant inheritance of the disease phenotype with complete penetrance (Fig. 1a). The mutant allele was not detected in unaffected family members or in 1,136 unrelated subjects of diverse ethnicity (Supplementary Fig. 3), making it unlikely that R1108X is a rare polymorphism. In the proband (the index case), sequencing of 100 additional regulatory genes essential for, or expressed during, cardiac development identified no other linked mutations, consistent with a monogenic aetiology (V.G. and D.S., unpublished observations).

Direct sequencing of *NOTCH1* in a smaller, unrelated Hispanic family with aortic valve disease revealed a second mutation that segregated with three affected family members, all with bicuspid aortic valve (Fig. 1d, e). The proband (III-1) also had mitral valve atresia, hypoplastic left ventricle and double-outlet right ventricle; his sibling (III-2) and mother (II-1) had aortic valve calcification and stenosis. Family member II-2 had an ascending aortic aneurysm (Fig. 1e; see also Supplementary Fig. 1b) but no aortic valve disease. A single base pair deletion at position 4515 that segregated with aortic valve disease in this family was not found in 1,138 ethnically diverse controls (Fig. 1f; see also Supplementary Fig. 3). This deletion resulted in a frameshift mutation (H1505del) that predicted a severely altered protein containing 74 incorrect amino acids at the carboxy terminus of the extracellular domain followed by a premature stop codon (Fig. 1f). These *NOTCH1* mutations generate truncated transcripts that probably undergo nonsense-mediated decay⁸ and provide compelling genetic evidence that *NOTCH1* haploinsufficiency results in human congenital heart disease, although dominant-negative effects cannot formally be ruled out.

To determine whether *Notch1* expression during development correlates with the predominant phenotype in humans, we performed *in situ* hybridization at multiple stages during cardiogenesis, focusing on *Notch1* transcripts in cardiac valves and their precursors in mice. At mouse embryonic day (E) 11.5, *Notch1* messenger RNA transcripts were abundant in the outflow tract mesenchyme, which gives rise to the valves, and in the endocardium (Fig. 2a–d). By E13.5,

when septation of the common arterial trunk occurs, *Notch1* was expressed at high levels in the endothelial layer (Fig. 2e–h) and mesenchyme of aortic valve leaflets (Fig. 2g, h), possibly explaining the increased dose sensitivity of the aortic valve. These findings suggest a role for Notch signalling in the morphological development of the aortic valve. Disruption of *Notch1* in mice results in death by E9.5 from vascular endothelial defects, precluding analysis of the aortic valve⁹, but in fish and frogs *Notch1* appears to be important for early valve development¹⁰.

NOTCH1 encodes a large protein containing an extracellular domain with 36 tandem epidermal growth factor (EGF)-like repeats and three cysteine-rich Notch/LIN-12 repeats, an intracellular domain with six ankyrin repeats, and a transactivation domain. The Notch signalling pathway is highly conserved across species⁷. Notch receptors (1–4) interact with Delta(1–4) or jagged(1, 2)/serrate, resulting in two independent cleavages, first by a metalloprotease^{11,12} and then by presenilin^{13,14}, that release the Notch

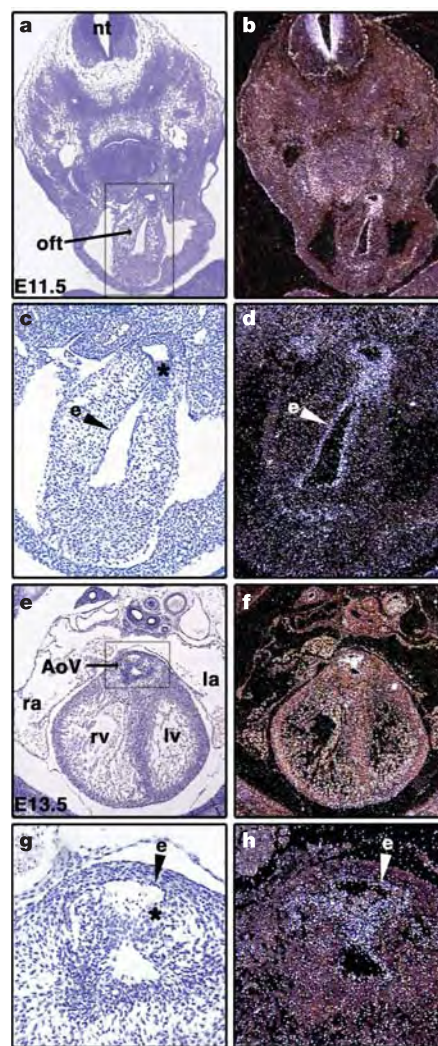


Figure 2 | Cardiac expression of mouse *Notch1* mRNA by radioactive-section *in situ* hybridization. **a–d**, Transverse sections of E11.5 mouse embryos through endocardium (arrowhead) and endocardial cushions (asterisk) of the outflow tract (oft). **c, d**, High-magnification images of **a**, **b**, as depicted by box in **a**. **e**, endocardium; **nt**, neural tube. **e–h**, Transverse sections through E13.5 embryonic heart. **g, h**, High-magnification of aortic valve (AoV) region outlined by box in **e**. The asterisk indicates AoV mesenchyme. Sections in **a**, **c**, **e** and **g** are bright-field images of **b**, **d**, **f** and **h**, respectively. **la**, left atrium; **lv**, left ventricle; **ra**, right atrium; **rv**, right ventricle.

intracellular domain from the membrane, in a manner similar to that first described for sterol response element binding protein¹⁵. Notch intracellular domain translocates to the nucleus, where it interacts with the DNA-binding protein CSL (CBF-1, suppressor of hairless, and Lag-1) to activate downstream target genes, including members of the hairy/enhancer of split (Hes) family of transcriptional repressors. This pathway participates in cell fate determination and differentiation during organogenesis throughout the embryo and is regulated by glycosylation of the extracellular EGF-like repeats in Notch1 (ref. 16).

Although haemodynamic alterations induced by bicuspid aortic valve may contribute to calcification, several family members with tricuspid aortic valves also developed calcification. This observation, along with the severity of calcification, led us to test whether NOTCH1 may directly affect calcium deposition. A proposed cellular mechanism by which valvular calcification develops is via differentiation of valvular cells into osteoblast-like cells⁶, including upregulation of genes, such as osteopontin¹⁷. Expression of osteopontin, osteocalcin and other osteoblast-specific genes is directly regulated by upstream *cis* elements that bind to the transcription factor Runx2 (ref. 18). Because Runx2 is upregulated in mouse and rabbit models of valvular calcification^{19,20}, we investigated whether Notch1 normally represses Runx2 activation. In the fibroblast cell line COS7, constitutively active Notch1 intracellular domain repressed Runx2-induced activation of luciferase through a multi-merized Runx2-binding *cis* element normally present upstream of osteocalcin (Fig. 3a). Although Runx2 can be inhibited by histone deacetylase (HDAC) activity²¹, Notch1-mediated repression of Runx2 was unaffected by the potent HDAC inhibitor trichostatin A, suggesting that it was HDAC-independent.

Notch directly activates the hairy family of transcriptional repressors, the central mediators of Notch's effects on gene expression. The heart and vasculature are enriched in the hairy-related transcriptional repressors Hrt1 and Hrt2, which mediate the Notch signal^{22,23}.

Hrt1 and Hrt2, also known as Hey1 and Hey2, were co-expressed in the endothelial lining of the murine aortic valve leaflet at E17.5, as well as in the endocardium and vascular endothelium (Fig. 3b–d). As with Notch1, Hrt1 and Hrt2 inhibited Runx2 activation of the osteocalcin enhancer. Using Hrt2 truncation mutants, we determined that the basic helix–loop–helix (bHLH) domain of Hrt2 is necessary for full Hrt2-mediated repression of Runx2 (Fig. 3e, f). Moreover, semi-quantitative RT–PCR demonstrated upregulation of Hrt1 and Hrt2 transcripts in COS7 cells transfected with Notch1 intracellular domain (Fig. 3g). In glutathione *S*-transferase (GST) pull-down studies performed to investigate the mechanism of Hrt2-mediated repression, Hrt2 and Runx2 specifically interacted (Fig. 3h), consistent with the repression we observed and the reported *in vitro* interaction between Hes1 and Runx2 (ref. 24). The bHLH domain of Hrt2 can recruit HDACs²⁵; however, as with Notch1, Hrt repression of Runx2 activation was not dependent on HDAC activity (Fig. 3). These data suggest that Hrt proteins repress Runx2 through a physical interaction, and may mediate Notch1 repression of Runx2.

Somatic *NOTCH1* mutations have been identified in human blood cancers²⁶, but the discovery of *NOTCH1* mutations as a cause of aortic valve calcification and aortic valve anomalies represents the first demonstration of *NOTCH1* germline mutations in human disease. The families reported here provided insights into the cause of a common human developmental malformation (bicuspid aortic valve) and revealed a potential mechanism mediated by NOTCH1 mutations that may predispose to endothelial dysfunction and inflammation underlying abnormal cardiovascular calcification events. Further studies of the NOTCH1 signalling pathway in the adult calcific process may identify preventative and pharmacological approaches to slow this age-related disease.

Whereas the role of NOTCH1 in preventing aortic valve calcification is relevant for adult-onset disease, its essential function in normal development of the valve is equally intriguing. Bicuspid

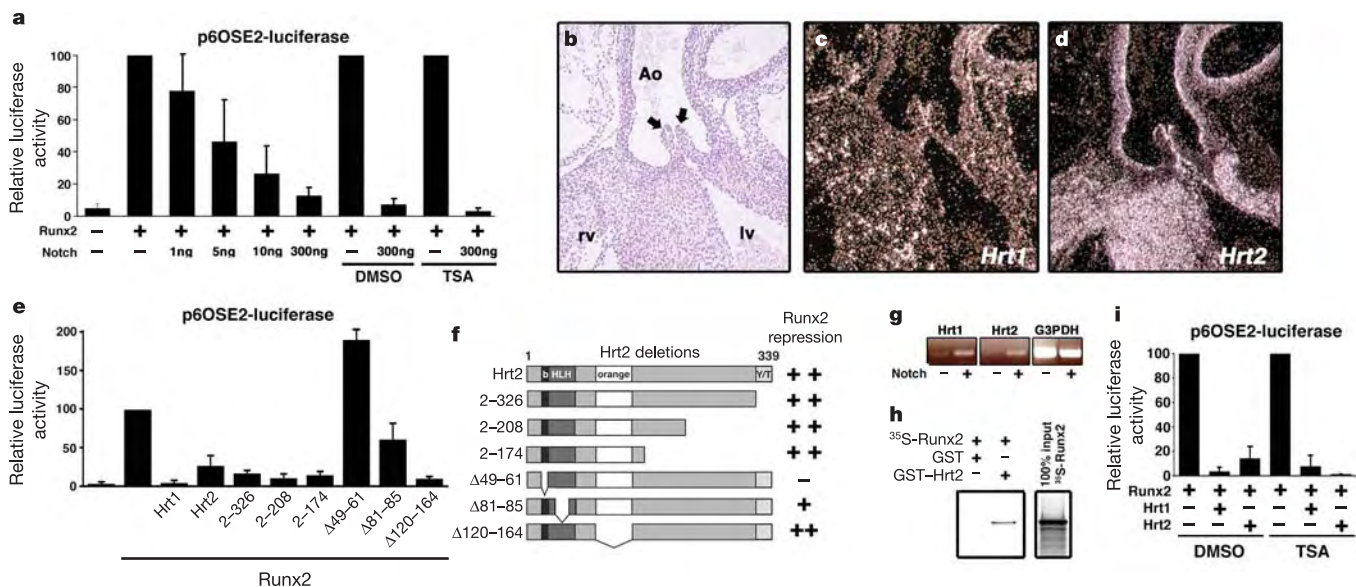


Figure 3 | Notch1, Hrt1 and Hrt2 repress Runx2 transcriptional activity. **a**, Relative luciferase activity in COS7 cells transfected with Flag-Runx2 and Runx2-dependent osteocalcin enhancer (p6OSE2) luciferase reporter with or without co-transfection of the indicated concentrations of Myc-Notch1 intracellular domain, in the presence of trichostatin A (TSA) or vehicle (DMSO). **b–d**, Coronal section *in situ* hybridization through aortic valve (arrows) of E17.5 mouse embryos (**c**, **d**). Panel **b** is a bright-field image. Ao, aorta; lv, left ventricle; rv, right ventricle. **e**, Relative luciferase activity directed by Runx2 with co-transfection of Hrt1, Hrt2 or Hrt2 mutant proteins. Numbers indicate amino acid positions of mutant proteins.

f, Schematic and summary of Hrt2 mutant protein effects on Runx2. **b**, basic domain; HLH, helix–loop–helix domain; orange, orange domain; Y/T, YXPW–TEIGAF motif. **g**, Semi-quantitative RT–PCR of Hrt1, Hrt2 and G3PDH in COS7 cells with or without transfection of Notch1 intracellular domain. **h**, Pull-down assays with GST–Hrt2 fusion protein and ³⁵S-labelled Runx2. **i**, Relative luciferase activity of Runx2 with Hrt1 or Hrt2 in the presence of trichostatin A or DMSO control. Luciferase data are shown as percentage of Runx2 activation (normalized to 100%); mean $\pm$ s.d. are shown.

and even unicuspid aortic valves typically contain a ridge where the valve leaflets did not separate *in utero* (Fig. 1g). In extreme cases, blood flow may be so restricted that the left ventricle fails to grow, resulting in hypoplastic left heart syndrome, the most frequent cause of death in children with congenital heart disease. As bicuspid aortic valve and hypoplastic left heart syndrome may represent extremes of the aortic valve disease spectrum, the discovery of *NOTCH1* as a cause of bicuspid aortic valve and a hypoplastic left ventricle in the same family suggests that *NOTCH1* mutations may be the genetic basis for hypoplastic left heart syndrome in some patients. Future studies of *NOTCH1* mutations in this population may reveal those at risk for a subset of severe congenital heart lesions.

METHODS

Clinical phenotype evaluation and DNA collection. The congenital heart disease families and individuals were ascertained for genetic linkage analyses at Children's Medical Center, Dallas (University of Texas Southwestern Medical Center) and the University of California, San Diego. Clinical evaluations and genetic studies were performed in accordance with human subject guidelines after informed consent according to the protocol approved by the individual Institutional Review Boards. Family members were studied by history, physical examination, 12-lead electrocardiogram, and echocardiography. Medical records were reviewed for individuals who had died. All phenotypic information was reviewed by cardiologists. Genomic DNA for genetic analyses was extracted from peripheral lymphocytes.

Genetic linkage analysis. Autosomal genome linkage analysis was performed with 372 polymorphic DNA markers at ~10-cM intervals (ABI Mapping Set v2.5). Markers were genotyped in all family members, and linkage analysis was performed with GENEHUNTER as described^{27,28}. In brief, initial linkage analysis of family A, assuming 90% penetrance and a disease allele frequency of 1.5%, demonstrated the highest LOD score on chromosome 9q34-35. Phenotypic analysis assuming 100% penetrance yielded a single peak at 9q34-35 and a maximum LOD score of 3.5.

Identification of *NOTCH1* mutations. All *NOTCH1* exons were sequenced bidirectionally to search for sequence variations in the probands of families A and B. Exons containing R1108X and H1505del mutations were amplified by PCR for each additional family member and sequenced bidirectionally. Sequences of the 42 primer pairs for the 34 *NOTCH1* exons are available on request. PCR amplification was performed with the BD Biosciences Advantage GC Genomic PCR kit following the manufacturer's instructions, with annealing at 60°C. Screening of identified human *NOTCH1* mutations was performed with allelic discrimination assays and the ABI Prism 7900 HT Sequence Detection System using TaqMan probes on DNA from participants in the Dallas Heart Study, as described²⁹.

Radioactive-section *in situ* hybridization. ³⁵S-labelled antisense riboprobes were synthesized with T7 RNA polymerase (MAXIScript, Ambion) from 400-bp partial mouse Notch1 cDNA or plasmids encoding Hrt1 and Hrt2. With these riboprobes, radioactive-section *in situ* hybridization was performed on paraffin-embedded sections of E11.5, E13.5 and E17.5 mouse embryos, as described²².

Luciferase assays. COS7 cells were transfected using Eugene 6 (Roche) according to the manufacturer's instructions. The reporter plasmid (250 ng), pOSE2 luciferase¹⁸, and CMV β-galactosidase expression plasmid (50 ng) to control for transfection efficiency were transfected along with Runx2 expression plasmid (100 ng) and Hrt1, Hrt2, Notch1 intracellular domain and Hrt2 deletion expression plasmids (300–1,000 ng). Hrt2 deletion constructs were generated as described³⁰ and protein levels of mutants kept constant. To inhibit HDAC activity, trichostatin A diluted to 0.1 μM in dimethyl sulphoxide (DMSO) was added 24 h before collecting cell lysates. Simultaneous duplicate experiments with an identical amount of DMSO served as a control. Immunoblots verified appropriate protein expression. Luciferase activity was measured 40 h after transient transfection as described³⁰ and was normalized to LacZ expression to generate relative luciferase activity (Fig. 3a) or expression of Hrt and Runx2 proteins (Fig. 3e). At least three independent experiments were performed in duplicate.

Quantitative RT-PCR. Total RNA from COS7 cells collected 40 h after transfection with empty vector and Notch1 intracellular domain expression plasmid was purified with the Trizol method (Invitrogen). Total RNA (1 μg) was reverse transcribed using the Superscript First Strand Synthesis System for RT-PCR (Invitrogen). PCR analysis was performed using primers specific for Hrt1 and Hrt2. Glyceraldehyde 3-phosphate dehydrogenase (G3PDH) RNA was amplified as a loading control. An annealing temperature of 56°C was used for PCR analysis. Negative controls for each sample used non-reverse-transcribed RNA.

GST pull-down assay. Mouse GST-Hrt2 fusion proteins were purified with glutathione Sepharose 4 Fast Flow beads (Roche) for 12 h and washed twice in binding buffer (300 mM NaCl, 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 1% Triton X-100, 1% NP-40, 0.1% SDS, 0.5 mM dithiothreitol). ³⁵S-labelled Runx2 protein was synthesized using the T7 TNT coupled reticulocyte lysate system (Promega) according to the manufacturer's instructions. Labelled protein was incubated with GST fusion protein (2 μg) for 8–10 h at 4°C in binding buffer with 1 mg of nonfat dried milk to compete for nonspecific interactions. Bound proteins were analysed by SDS-PAGE and autoradiography.

Received 2 May; accepted 17 June 2005.

Published online 17 July 2005.

1. American Heart Association, *Heart Disease and Stroke Statistics—2004 Update* (American Heart Association, Dallas, Texas, 2004).
2. Hoffman, J. I. & Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39**, 1890–1900 (2002).
3. Cripe, L., Andelfinger, G., Martin, L. J., Shooner, K. & Benson, D. W. Bicuspid aortic valve is heritable. *J. Am. Coll. Cardiol.* **44**, 138–143 (2004).
4. Horne, B. D., Camp, N. J., Muhlestein, J. B. & Cannon-Albright, L. A. Evidence for a heritable component in death resulting from aortic and mitral valve diseases. *Circulation* **110**, 3143–3148 (2004).
5. Loffredo, C. A. *et al.* Prevalence of congenital cardiovascular malformations among relatives of infants with hypoplastic left heart, coarctation of the aorta, and d-transposition of the great arteries. *Am. J. Med. Genet.* **124A**, 225–230 (2004).
6. Rajamannan, N. M., Gersh, B. & Bonow, R. O. Calcific aortic stenosis: from bench to the bedside—emerging clinical and cellular concepts. *Heart* **89**, 801–805 (2003).
7. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signalling: cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
8. Frischmeyer, P. A. *et al.* An mRNA surveillance mechanism that eliminates transcripts lacking termination codons. *Science* **295**, 2258–2261 (2002).
9. Krebs, L. T. *et al.* Notch signalling is essential for vascular morphogenesis in mice. *Genes Dev.* **14**, 1343–1352 (2000).
10. Timmerman, L. A. *et al.* Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. *Genes Dev.* **18**, 99–115 (2004).
11. Wen, C., Metzstein, M. M. & Greenwald, I. SUP-17, a *Caenorhabditis elegans* ADAM protein related to *Drosophila* KUZBANIAN, and its role in LIN-12/NOTCH signalling. *Development* **124**, 4759–4767 (1997).
12. Sotillos, S., Roch, F. & Campuzano, S. The metalloprotease-disintegrin Kuzbanian participates in Notch activation during growth and patterning of *Drosophila* imaginal discs. *Development* **124**, 4769–4779 (1997).
13. Struhl, G. & Adachi, A. Nuclear access and action of notch *in vivo*. *Cell* **93**, 649–660 (1998).
14. Struhl, G. & Greenwald, I. Presenilin is required for activity and nuclear access of Notch in *Drosophila*. *Nature* **398**, 522–525 (1999).
15. Brown, M. S. & Goldstein, J. L. The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. *Cell* **89**, 331–340 (1997).
16. Haines, N. & Irvine, K. D. Glycosylation regulates Notch signalling. *Nature Rev. Mol. Cell Biol.* **4**, 786–797 (2003).
17. O'Brien, K. D. *et al.* Osteopontin is expressed in human aortic valvular lesions. *Circulation* **92**, 2163–2168 (1995).
18. Dacy, P., Zhang, R., Geoffroy, V., Ridall, A. L. & Karsenty, G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell* **89**, 747–754 (1997).
19. Steitz, S. A. *et al.* Smooth muscle cell phenotypic transition associated with calcification: upregulation of Cbfa1 and downregulation of smooth muscle lineage markers. *Circ. Res.* **89**, 1147–1154 (2001).
20. Rajamannan, N. M. *et al.* Atorvastatin inhibits hypercholesterolemia-induced cellular proliferation and bone matrix production in the rabbit aortic valve. *Circulation* **105**, 2660–2665 (2002).
21. Vega, R. B. *et al.* Histone deacetylase 4 controls chondrocyte hypertrophy during skeletogenesis. *Cell* **119**, 555–566 (2004).
22. Nakagawa, O., Nakagawa, M., Richardson, J. A., Olson, E. N. & Srivastava, D. HRT1, HRT2, and HRT3: a new subclass of bHLH transcription factors marking specific cardiac, somitic, and pharyngeal arch segments. *Dev. Biol.* **216**, 72–84 (1999).
23. Nakagawa, O. *et al.* Members of the HRT family of basic helix-loop-helix proteins act as transcriptional repressors downstream of Notch signalling. *Proc. Natl Acad. Sci. USA* **97**, 13655–13660 (2000).
24. McLaren, K. W. *et al.* The mammalian basic helix loop helix protein HES-1 binds to and modulates the transactivating function of the runt-related factor Cbfa1. *J. Biol. Chem.* **275**, 530–538 (2000).
25. Iso, T. *et al.* HERP, a novel heterodimer partner of HES/E(spl) in Notch signalling. *Mol. Cell Biol.* **21**, 6080–6089 (2001).
26. Weng, A. P. *et al.* Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science* **306**, 269–271 (2004).

27. Garg, V. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**, 443–447 (2003).
28. Kruglyak, L., Daly, M. J., Reeve-Daly, M. P. & Lander, E. S. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
29. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature Genet.* **37**, 161–165 (2005).
30. Kathiriya, I. S. *et al.* Hairy-related transcription factors inhibit GATA-dependent cardiac gene expression through a signal-responsive mechanism. *J. Biol. Chem.* **279**, 54937–54943 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors thank the families for their participation; the Divisions of Pediatric Cardiology and Pediatric Cardiothoracic Surgery at Children's Medical Center Dallas for assistance with clinical information; McDermott Center for Human Growth and Development for assistance with linkage analysis and allelic discrimination assays; Dallas Heart Study participants

and investigators for DNA samples; members of the Molecular Histology Core laboratory for radioactive-section *in situ* hybridization; J. C. Cohen and H. H. Hobbs for discussions and critical review of this manuscript; K. Ivey for graphics assistance; C. Butler, A. Garg and D. Srivastava for assistance with blood collection; G. Karsenty for Runx2 expression and p6OSE2 reporter plasmids; and L. Kedes for the GST-Hrt2 expression plasmid. This work was supported by grants from NICHD/NIH and March of Dimes Birth Defects Foundation to V.G., and NHLBI/NIH, March of Dimes Birth Defects Foundation and the Donald W. Reynolds Cardiovascular Clinical Research Center to D.S. J.F.R. was supported by a training grant from NIH; I.N.K. is an NICHD/NIH fellow of the Pediatric Scientist Development Program; and D.S. is an Established Investigator of the American Heart Association.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.G. (vidu.garg@utsouthwestern.edu) or D.S. (dsrivastava@gladstone.ucsf.edu).

Asymmetric cell divisions promote stratification and differentiation of mammalian skin

Terry Lechler¹ & Elaine Fuchs¹

The epidermis is a stratified squamous epithelium forming the barrier that excludes harmful microbes and retains body fluids. To perform these functions, proliferative basal cells in the innermost layer periodically detach from an underlying basement membrane of extracellular matrix, move outward and eventually die. Once suprabasal, cells stop dividing and enter a differentiation programme to form the barrier¹. The mechanism of stratification is poorly understood. Although studies *in vitro* have led to the view that stratification occurs through the delamination and subsequent movement of epidermal cells^{2–4}, most culture conditions favour keratinocytes that lack the polarity and cuboidal morphology of basal keratinocytes in tissue. These features could be important in considering an alternative mechanism, that stratification occurs through asymmetric cell divisions in which the mitotic spindle orients perpendicularly to the basement membrane^{5–7}. Here we show that basal epidermal cells use their polarity to divide asymmetrically, generating a committed suprabasal cell and a proliferative basal cell. We further demonstrate that integrins and cadherins are essential for the apical localization of atypical protein kinase C, the Par3–LGN–Inscuteable complex and NuMA–dynactin to align the spindle.

We first addressed whether oriented cell divisions participate in stratification during epidermal development. At embryonic day 12.5 (E12.5), most of the epidermis was single-layered, and most divisions occurred laterally, within the plane of the epithelium. Unexpectedly, however, a few mitotic cells seemed to be dividing perpendicularly to the basement membrane, as judged by staining with 4',6-diamidino-2-phenylindole (DAPI) and with anti-tubulin (arrow in Fig. 1a, b). Closer inspection revealed the presence of some suprabasal cells within the E12.5 single-layered epithelium. In these regions, the occasional suprabasal cell was often positioned directly over a basal cell, as would be expected for a perpendicular division.

To facilitate quantification, we engineered mice expressing keratin 14 (K14)–centrin coupled to green fluorescent protein (GFP) and examined embryos at later stages of development, as the epidermis became multi-layered. It was now easy to distinguish parallel from perpendicular spindle alignments (Fig. 1c, d). In predominantly single-layered areas of E12.5 epithelium, 92% of divisions occurred parallel to the basement membrane. By contrast, in E12.5 areas that showed early signs of stratification, most mitotic cells had an alignment that was clearly perpendicular. At E12.5, perpendicularly aligned spindles accounted for about 22% of all mitoses, but as stratification progressed across the epithelium, this number rose markedly (Fig. 1e). From E15.5 onwards through postnatal development, more than 70% of spindles were oriented perpendicularly to the basement membrane.

Perpendicular alignments yielded one basal and one suprabasal cell, and their temporal appearance during skin development correlated well with stratification. A functional link was provided by examining *p63*-null embryos, defective in the transcription factor

p63 but also in stratification^{8,9}. Divisions in E18.5 *p63*-null epidermis were lateral to the basement membrane (an example is shown in Fig. 1f). In normal adults, homeostasis of stratification also seemed to be dependent on perpendicular divisions, because most divisions

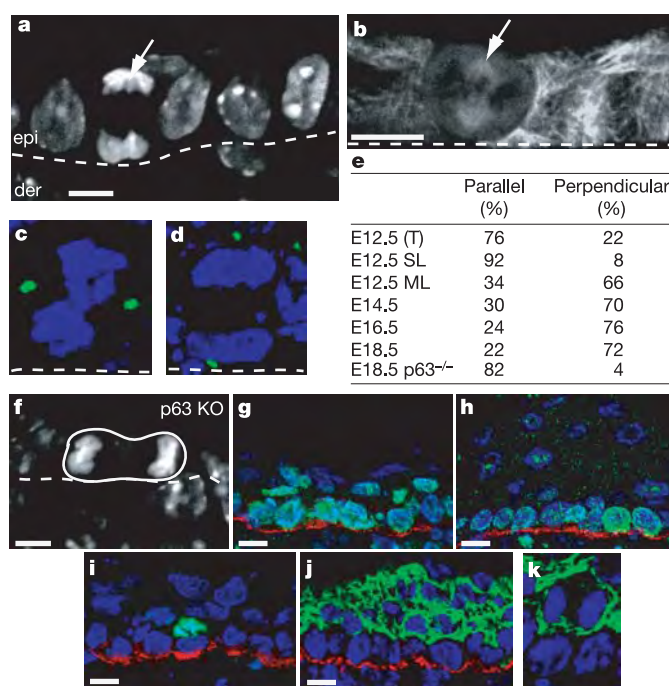


Figure 1 | Asymmetric cell divisions govern stratification and differentiation during epidermal development. **a–d**, Images of embryonic skin showing the orientation of mitoses relative to the basement membrane (white dotted line), separating epidermis (epi in **a**) from dermis (der). Arrows indicate the mitotic cells in E12.5 epidermis; **a** (anaphase) and **b** (metaphase). DAPI marks the DNA in **a**. Indirect immunofluorescence with tubulin antibodies marks the spindle in **b**. Transgenic Centrin-GFP (green) marks spindle poles in **c** (metaphase, parallel spindle orientation), and **d** (anaphase, perpendicular spindle orientation). **e**, Quantification of division planes in epidermis during development. SL, single-layered E12.5 epidermis, ML, multi-layered E12.5 epidermis, T, total E12.5 epidermis. **f**, Parallel spindle orientation of an anaphase cell in *p63* KO epidermis at E18.5; DAPI staining. **g, h**, Antibodies against $\beta 4$ integrin (red) label the base of the basal cells. Antibodies against Ki67 (green) show that not all suprabasal cells at E15.5 (**g**) have withdrawn from a proliferative state; however, Ki67 is confined to basal cells at E18.5 (**h**). **i**, Antibodies against phospho-histone H3 (green) reveal the presence of mitotic suprabasal keratinocytes at E15.5. Red staining shows $\beta 4$ integrin. **j**, Keratin 1 immunofluorescence (green) shows that suprabasal cells have turned on this differentiation marker at E15.5. Red staining shows $\beta 4$ integrin. **k**, Magnified view of a keratin-1-positive (green) suprabasal mitotic cell in anaphase. Scale bars are 10 μ m.

¹Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10021, USA.

in both adult epidermis (about 85%) and tongue (about 65%) still occurred in this fashion, even though overall mitoses waned considerably once animals reached full size.

A feature of these unexpected perpendicular spindle orientations is that one cell maintains and the other loses contact with the underlying basement membrane. Many key functions including extracellular matrix protein secretion, integrin-mediated focal adhesion and growth factor signalling are already known to be basally localized, and perpendicular divisions provide a natural mechanism for their unequal partitioning to two daughter cells^{1,10,11}. Perpendicular spindle orientations in the basal epidermal layer therefore automatically result not only in stratification but also in asymmetric cell divisions.

Using K14–histone H2B mice to label the epidermis uniformly¹², we determined that the population of basal cells increased about threefold between E13.5 and E15.5. This expansion reflected the increasing size of the embryo and the ability of basal cells to divide symmetrically and remain attached to the basement membrane. By contrast, the increase in suprabasal cells during this time was much too large to be explained by the concomitant increase in asymmetric cell divisions (see Fig. 1e). This conundrum was solved by discovering that, in contrast to fully stratified E18.5 and postnatal epidermis,

many E15.5 suprabasal cells were proliferative. They expressed the proliferation marker Ki67, and in mitoses they possessed phosphorylated histone H3 (Fig. 1g–i). E15.5 suprabasal cells, including mitotic cells, were nevertheless positive for the differentiation marker keratin-1 (Fig. 1j, k). The mitotic potential of these cells was shown by using monastrol or nocodazole to trigger the spindle checkpoint, arresting 15–20% of suprabasal cells in mitosis within 5 h (data not shown). By retaining mitotic activity at this age, the suprabasal population could expand more rapidly to form the multilayered epidermis. Nevertheless, the proliferative potential of suprabasal cells was short-lived and they were positive for differentiation markers.

How do epidermal cells control their spindle orientation and divide asymmetrically? In *Drosophila* neuroblasts, a protein complex containing Inscuteable and Partner of Inscuteable (Pins) captures a spindle pole at the apical cortex, aligning the spindle with the apical–basal axis^{7,13–16}. Although mammalian LGN is a genuine Pins homologue, no Inscuteable counterpart has been cloned. Reports of asymmetric divisions in mammals are rare, and it remains unknown whether this mechanism has been conserved.

We therefore cloned a distant mouse homologue of *Drosophila* Inscuteable, naming it mInsc. Expression of the *mInsc* gene in embryonic, newborn and adult skin was confirmed by northern

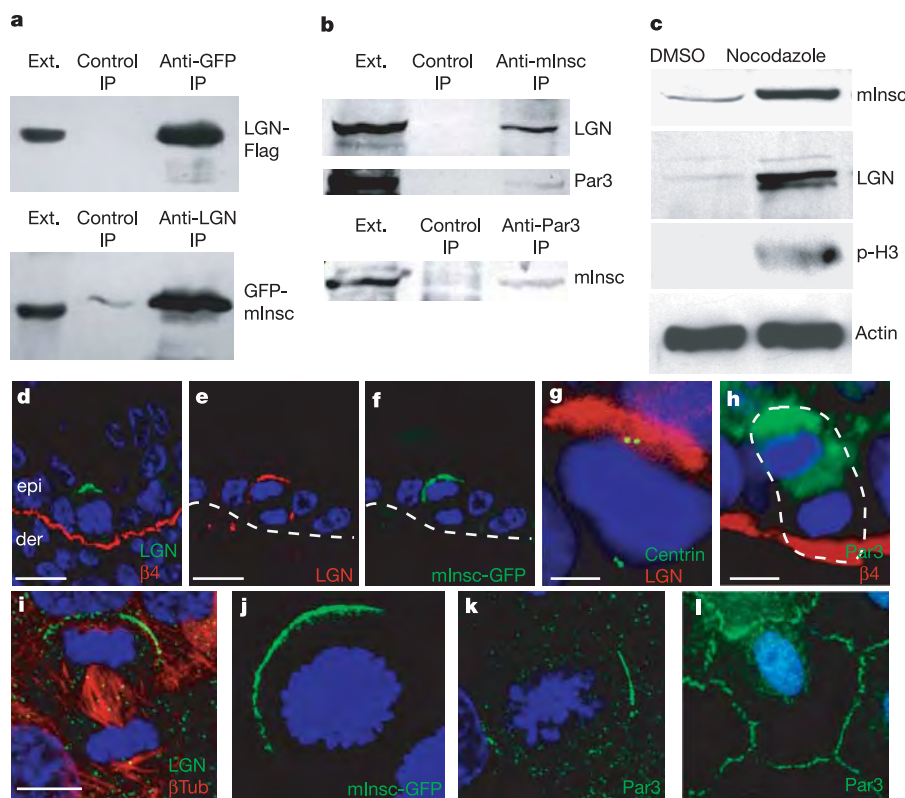


Figure 2 | Mitotic apical localization of a mInsc-LGN-Par3 complex.

a, mInsc interacts with LGN in cultured cells. Keratinocytes were transfected with K14 promoter driven full-length LGN (LGN-Flag) and mInsc (GFP-mInsc). Reciprocal co-immunoprecipitations (IPs) were performed. The extract (Ext) lane contains 10% of the input protein for the co-immunoprecipitation. **b**, Endogenous mInsc, LGN and Par3 form a complex in epidermis from E14.5 embryos. Protein extracts were prepared from embryos treated with nocodazole for 5 h to enrich for mitotic cells. Extracts were immunoprecipitated with anti-mInsc or anti-Par3 antibody as indicated, and immunoblots of the IPs were then performed with LGN, Par3 and mInsc antibodies. **c**, Epidermal protein lysates were prepared from E14.5 embryos treated with dimethylsulphoxide (DMSO; vehicle) or nocodazole for 5 h and probed with the indicated antibodies. **d–l**, Epifluorescence (GFP) or indirect immunofluorescence with antibodies colour-coded as indicated. DAPI (blue) was used to label chromatin.

d–f, Apical crescents of LGN (**d**, **e**) and mInsc-GFP (**f**) are present in most mitotic basal cells of E15.5 embryonic epidermis (**d**) and adult (**e**, **f**). epi, epidermis; der, dermis. Note that **e** and **f** show co-localization of LGN and mInsc-GFP to the apical cortex of an anaphase cell. **g**, One E14.5 spindle pole was always located directly underneath the LGN crescent. **h**, Par3 localizes at E15.5 to the apical cortex of mitotic basal keratinocytes. The white dashed line demarcates basement membrane in **e** and **f** and borders of the anaphase cell in **h**. **i**, Mouse skin keratinocytes were stained with anti-LGN and β -tubulin antibodies. **j**, Keratinocytes were transfected with K14-GFP-mInsc and labelled simultaneously with DAPI to localize the DNA and identify mitotic cells. The image is a compressed Z-stack of confocal sections to view the entire cell. **k**, **l**, Keratinocytes were stained for DAPI and Par3 in mitotic (**k**) and interphase (**l**) cells. Scale bars are 10 μ m, except in **g**, **h**, in which they are 5 μ m.

blot and reverse transcriptase polymerase chain reaction analyses (Supplementary Fig. S1a, b). Because mInsc and Inscuteable share only about 20% sequence identity (Supplementary Fig. S1c), we first examined whether mInsc interacts with LGN and Par3, the two known binding partners for Inscuteable. When transiently expressed together in cultured keratinocytes, epitope-tagged mInsc and LGN formed a complex that could be immunoprecipitated by either anti-mInsc or anti-LGN (Fig. 2a). The endogenous mInsc, LGN and Par3 proteins also existed as a complex. Thus, when extracts of E14.5 skin epidermis were immunoprecipitated with a newly raised anti-mInsc antibody, immunoblot analyses detected the three proteins in the complex (Fig. 2b). Similarly, immunoprecipitates of Par3 contained mInsc. The total levels of these proteins were always significantly greater in embryonic skin enriched for mitotic keratinocytes (Fig. 2c). This coordinate mitotic regulation of mInsc and LGN is consistent with their putative role in governing spindle alignment.

Immunofluorescence microscopy revealed further insights into the role of LGN–mInsc complexes in developing epidermis. Cells in interphase or undergoing lateral divisions, either in single-layered or multilayered epidermis, had a diffuse localization of LGN (data not shown). From E15.5 onwards, however, an apical crescent of anti-LGN labelling was seen at the cortex of most mitotic cells (Fig. 2d). Closer inspection revealed that in about 90% of cells with an apical crescent of LGN, one of the spindle poles was positioned directly below it, indicative of a perpendicular alignment of the spindle (Fig. 2g; adult). Co-localization of mInsc and LGN was best revealed *in vivo* by engineering a transgenic mouse expressing mInsc-GFP in the epidermis. An apical crescent of mInsc-GFP, LGN and Par3 was seen in mitotic basal cells within the stratified epidermis (examples are shown in Fig. 2e–h). Thus, a strict correlation existed between asymmetric cell divisions, stratification and the crescent-shaped apical organization of LGN, mInsc and Par3. Moreover, the localization of the LGN–mInsc–Par3 crescent was always opposite to that of the integrins and basement membrane (Fig. 2d, h).

LGN, mInsc-GFP and Par3 had a mitosis-specific polarized distribution in cultured cells, often localizing asymmetrically to the cortex near one of the two spindle poles (Fig. 2i–k). Although polarization *in vitro* was more variable, it was observed in at least 15% and sometimes as many as 40% of mitotic keratinocytes. In non-dividing cells, these proteins were distributed at sites of cell–cell contact (Fig. 2l) or were diffuse (data not shown). mInsc–Par3–LGN polarization was unexpected given the flatter morphology of keratinocytes *in vitro*. However, the distribution occurred under conditions where a rich extracellular matrix was deposited, and where intercellular adhesion was present. We return to these issues below.

LGN has recently been shown to interact directly with NuMA, a large coiled-coil protein that is nuclear in interphase and organizes the spindle poles during mitosis in cultured cells^{17–20}. This and a possible relation to a *Caenorhabditis elegans* spindle protein Lin-5 (ref. 21) led us to investigate whether NuMA participates in aligning the spindle poles with LGN–Inscuteable–Par3 complexes to govern asymmetric cell divisions. Both cultured keratinocytes and cells within the basal layer of embryonic epidermis possessed the expected nuclear localization of NuMA during interphase (not shown) and spindle-pole localization of NuMA during mitosis (Fig. 3a, b). However, asymmetrically dividing basal cells frequently exhibited more intense and expanded anti-NuMA staining at their apical surface (examples are shown in Fig. 3a, b).

The crescent-shaped apical distribution of NuMA resembled that of the LGN complex, and immunofluorescence microscopy of K14-GFP-LGN-transfected keratinocytes confirmed their colocalization (Fig. 3c). Neither NuMA nor LGN required microtubules for their localization to the cortex, as judged by treatment with nocodazole (Fig. 3d). As reported earlier²², however, NuMA's localization to the centrosome was microtubule-dependent.

Asymmetric forces pulling on microtubules at the cell cortex seem

to be responsible for spindle reorientation in asymmetrically dividing cells within the *C. elegans* embryo²³. Although dynein–dynactin motor complexes at the cortex are believed to provide a pulling force on astral microtubules, they have not been observed to be asymmetrically localized in *Drosophila* neuroblast or in one-cell *C. elegans* embryos. Interestingly, in mitotic keratinocytes, polarized cortical immunolocalization was observed with antibodies against p150glued, a subunit of the dynactin complex (Fig. 3e). Moreover, dynactin colocalized with NuMA both at spindles/spindle poles and in the polarized cortical crescent (Fig. 3f). These findings indicate that NuMA and dynein–dynactin-dependent pulling forces at the apical cortex might function in asymmetric divisions.

Asymmetric LGN–mInsc polarization in epidermis differed from all previous examples in that it occurred in cells attached to an underlying basement membrane, a process known to be dependent on $\beta 4$ integrin (essential for hemidesmosomes)^{24,25} and $\beta 1$ integrin (essential for focal adhesions and basement membrane assembly)^{26,27}. Using mouse genetics, we uncovered the functional importance of the basement membrane in LGN–mInsc polarization. Thus,

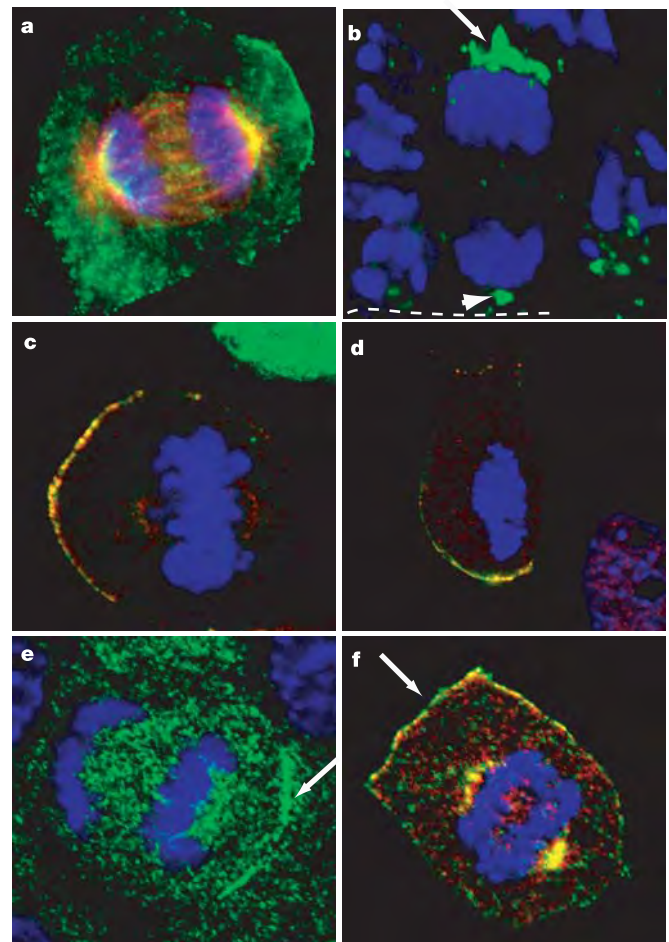


Figure 3 | Polarized mitotic localization of NuMA and dynactin. **a**, Mitotic keratinocyte stained simultaneously for NuMA (green) and microtubules (β -tubulin; red). **b**, E15.5 skin section stained with anti-NuMA antibodies reveals localization of NuMA (green) to spindle poles (arrowhead) and a cortical apical crescent (arrow). White dotted line denotes basement membrane. **c**, LGN (green) and NuMA (red) localize to the same crescent. K14-GFP-LGN-transfected keratinocytes were stained with antibodies against NuMA. **d**, Keratinocytes were treated with nocodazole to disrupt microtubules and then stained as in **c**. **e**, An anaphase keratinocyte stained for the p150glued subunit (green) of the dynactin complex. **f**, Co-localization of NuMA (green) and p150glued (red) in a mitotic keratinocyte. DNA is stained with DAPI (blue).

although proper apical complex localization was maintained in $\beta 4$ -null epidermis, it was aberrant in $\beta 1$ -null epidermis, which is unable to assemble a basement membrane (Fig. 4a, b). The LGN complex still formed a crescent in $\beta 1$ -null epidermis, but it was no longer restricted to the apical domain (Fig. 4b). Sometimes the crescent was even basal. The data were quantified and are presented in the diagrams in Fig. 4i, j. Each dot represents the centre of an LGN crescent in a mitotic cell.

The localization of NuMA was similarly aberrant in the $\beta 1$ -null epidermis (Fig. 4e). This was accompanied by abnormalities in spindle orientation and cell divisions. Figure 4m, n illustrates the spindle orientations measured in 25 mitotic cells. Taken together, these findings establish a role for $\beta 1$ and basement membrane in LGN complex distribution and in proper spindle orientation.

Next we examined the role of cell–cell adhesion in promoting LGN complex localization. Gene targeting in mice has revealed essential roles for desmoplakin (a component of desmosomes)²⁸ and α -catenin (a component of adherens junctions)²⁹ in intercellular adhesion within the epidermis. LGN complexes and NuMA had a normal localization in epidermis that lacked desmoplakin (Fig. 4c). However, in the absence of α -catenin LGN no longer localized to the cell cortex, with a resulting randomization of spindle alignment and misoriented cell divisions (Fig. 4d, o). These findings revealed an essential role for α -catenin in the targeting and/or maintenance of the polarized LGN crescent.

Although LGN (shown) and Par3 (not shown) lost their cortical association, NuMA showed merely a randomization of polarization in the α -catenin knockout (KO), more similar to $\beta 1$ -null epidermis

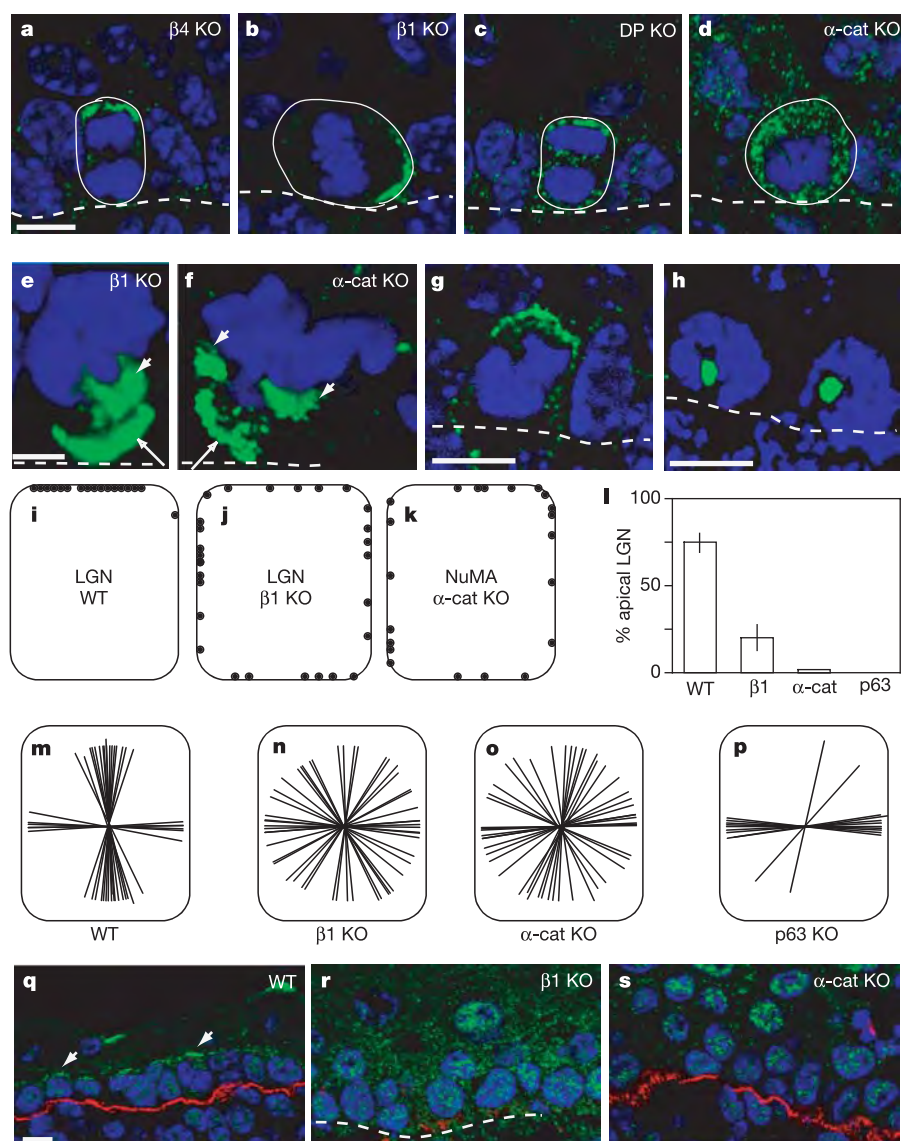


Figure 4 | Cadherin and integrin requirements for LGN–NuMA localization and spindle orientation. **a–d**, LGN localization (green staining) was determined in embryonic skin from knockout (KO) mice lacking $\beta 4$ integrin (**a**), $\beta 1$ integrin (**b**), desmoplakin (**c**) and α -catenin (**d**). **e, f**, NuMA localization (green) in $\beta 1$ -integrin KO (**e**) and α -catenin KO (**f**) epidermis. **g**, LGN (green) and **h**, NuMA (green) localization in sections from embryos treated with monastrol for 5 h. **i, j**, Schematic representation of LGN localization in WT (**i**) and $\beta 1$ KO (**j**) epidermis. Each dot represents the centre of an LGN crescent in a single mitotic cell. **k**, Schematic representation of NuMA localization in α -catenin KO epidermis.

l, Quantification of cells with apical cortical localization of LGN. Note that α -catenin (α -cat) and p63 KO do not show cortical localization of LGN. Error bars represent standard deviation. **m–p**, Schematic representation of spindle orientation in embryo skin from WT mice (**m**) and in $\beta 1$ KO (**n**), α -catenin KO (**o**) and p63 KO (**p**) mice. Each line represents the spindle axis of a single late metaphase or anaphase cell. **q–s**, Immunolocalization of PKC- ζ (green) in WT (**q**), $\beta 1$ -integrin KO (**r**) and α -catenin KO epidermis (**s**). Red staining shows $\beta 4$ integrin. Scale bars are 10 μ m, except in **e, f**, in which they are 5 μ m.

(Fig. 4f, k). However, in about 90% of cases, the spindle was aligned with the randomly positioned NuMA crescent. The loss of polarity and control of spindle orientation explains many of the architectural and biochemical defects in the α -catenin KO epidermis, including the absence of organized columns of stratified cells, the presence of suprabasal mitoses, the increased suprabasal inheritance and expression of integrins, the patchy keratin-1 expression and the formation of internalized masses of disorganized epidermal cells (see Supplementary Fig. S3 and ref. 29).

If LGN and NuMA interact directly, why do LGN complexes fail to polarize in the absence of α -catenin, whereas NuMA still forms a cortical crescent? Probing more deeply into this issue, we discovered that in wild-type (WT) epidermis, the kinesin Eg5 inhibitor monastrol resulted in the collapse of NuMA to the centre of the single microtubule monoastr, but LGN remained localized to the apical cortex (Fig. 4g, h). Taken together, these data indicate that the cortical localization of NuMA and LGN might be independent of each other, with LGN dependent on, and NuMA independent of, Par3 polarization.

Atypical protein kinase C (PKC- ζ) is part of a protein complex that is important for specifying the apical domain of polarized epithelia²⁹. Consistent with this notion was the observation that PKC- ζ was expressed in most if not all basal cells in WT epidermis, where it had a marked apical localization (Fig. 4q). This seemed to be independent of cell cycle status. However, PKC- ζ 's apical localization was lost in both β 1-null and α -catenin-null epidermis (Fig. 4r, s).

Finally, if apical localization of LGN–mInsc complexes is required for asymmetric cell divisions, then genetic models in which these divisions do not occur should lack this apical localization. We therefore turned again to the *p63*-null embryo. Unlike in their control littermates, apical accumulation of LGN was not detected in the *p63*-null epidermal keratinocytes (Fig. 4l). Concomitantly, and as shown in Fig. 4p, spindle orientations were overwhelmingly lateral. These data underscore the importance of the LGN–mInsc complexes in the governance of asymmetric cell divisions and in the organization and homeostasis of the epidermis as columnar units of stratified cells. The maintenance of lateral divisions in the *p63*-null epidermis is similar to that which occurs in simple epithelial tissues and contrasts with the random divisions that occur in the α -catenin-null epidermis. Although beyond the scope of the present study, the difference indicates that α -catenin might be involved in governing not only asymmetric divisions, dependent on LGN–mInsc, but also in symmetric divisions, which did not seem to depend on LGN–mInsc. This would be in agreement with genetic studies in *Drosophila*³⁰.

Few examples of asymmetric cell divisions have been documented in mammals, but we have shown that in epidermis, proper columnar stratification and tissue organization are driven at least in part by oriented, asymmetric cell divisions. In contrast to previous examples of asymmetric divisions, here the epidermis controls polarity by making intercellular connections with its neighbours as well as attaching to an underlying basement membrane. We found that α -catenin is essential not only for the apical recruitment of PKC- ζ but also for the assembly of the apical LGN–mInsc–Par3 crescent. Although the assemblies of the LGN–mInsc–Par3 and NuMA crescents are independent of one another, their apical localization and involvement in achieving the asymmetric cell division seem to be intimately linked. Given the role of α -catenin in the association of LGN with the cortex, it is noteworthy that adherens junctions (and focal adhesions) participate in orchestrating actin cytoarchitecture in epithelial tissues. It has been previously reported that disruption of F-actin causes a loss of cortical association of LGN³¹.

In addition, the basement membrane provides a natural mechanism for concentrating cell determinants at the base of the keratinocyte, and although β 1 integrin is not essential for the assembly of the Par3–LGN–mInsc–NuMA–dynactin crescent, it does have a function in apically orienting PKC- ζ and the crescent and hence dictating the directionality of the ensuing cell division. In so doing, this newly

formed apical crescent could become a molecular centre for the association of key cell commitment or differentiation factors, just as the basally oriented integrins are known to form a localization site for growth-promoting factors. In this regard, we showed that in the absence of p63, localization of the LGN–mInsc complex and perpendicular angling of the spindle plane do not occur and partitioning of determinants becomes symmetric, concomitant with impaired stratification and differentiation. Our findings now provide a framework for future studies in this area.

METHODS

Generation and immunoprecipitation of mInsc antibody. An antibody was raised in rabbits immunized with a peptide CGDKQRVDTPYTRDQ. By immunoblot analysis, affinity-purified antibody recognized a single band corresponding to the expected size of mInsc.

Protein extracts were prepared from E15.5 epidermis or cultured cells in buffer containing 50 mM Hepes pH 7.4, 100 mM KCl, 2 mM MgCl₂, 0.5% Triton X-100, supplemented with 1 mM dithiothreitol and protease inhibitors. Extracts were incubated with 2 μ g of control or specific antibody bound to protein G–Sepharose, then washed extensively with buffer; bound proteins were eluted with SDS sample buffer and subjected to immunoblot analysis. Other antibodies used included anti-LGN (S. Bahria), anti-Par3 (I. Macara), anti-NuMA (D. Compton), anti-Flag (Sigma), anti-GFP (Abcam), anti- β -tubulin (Sigma), anti-PKC- ζ (Santa Cruz), anti-Ki67 (Novocastro), anti-pH3 (Upstate) and anti-K1 (Fuchs laboratory).

Mitotic arrests were performed by incubating embryos in medium containing 10 μ M nocodazole or 100 μ M monastrol for 5 h at 37 °C.

Quantification of LGN–NuMA localization and spindle orientation. Late metaphase and anaphase spindles were used for quantification. Spindles that were oriented at $90 \pm 30^\circ$ to the basement membrane were classed as perpendicular; those that were oriented at $0 \pm 30^\circ$ were classed as parallel. LGN–NuMA apical localization was determined by dividing the cell into four quadrants, namely apical, basal and two lateral surfaces. We then determined the quadrant in which the centre of the LGN crescent lay. At embryonic stages, 100–200 mitoses were counted. In adults, because of the low mitotic index we found very few cells at a cell-cycle stage that could be quantified. For the epidermis we found 17 of 20 cells dividing asymmetrically, and 13 of 20 in the tongue. To determine the relationship between the NuMA cortical crescent and the spindle in α -catenin KO, we stained with NuMA and examined late metaphase and anaphase cells in which a clear cortical crescent, as well as one or more separated spindle poles was obvious. In 22 of 25 cases, the spindle was aligned with the cortical crescent.

Image acquisition. After fixation and staining, imaging was performed on a Zeiss Axioplan 2 microscope with Apotome attachment, using an AxioCam MRm camera, and Axiovision software. Quantification of basal cell numbers was performed by flow cytometry of cells isolated from full skin of K14–histone 2B mice labelled with β 4-integrin antibodies. Basal-to-suprabasal ratios were determined by manually counting basal (β 4-integrin-positive) and suprabasal (β 4-integrin-negative) cells in the microscope.

Received 18 March; accepted 8 June 2005.

Published online 10 August 2005.

1. Fuchs, E. & Raghavan, S. Getting under the skin of epidermal morphogenesis. *Nature Rev. Genet.* **3**, 199–209 (2002).
2. Vaezi, A., Bauer, C., Vasioukhin, V. & Fuchs, E. Actin cable dynamics and Rho/Rock orchestrate a polarized cytoskeletal architecture in the early steps of assembling a stratified epithelium. *Dev. Cell* **3**, 367–381 (2002).
3. Watt, F. M. Selective migration of terminally differentiating cells from the basal layer of cultured human epidermis. *J. Cell Biol.* **98**, 16–21 (1984).
4. Watt, F. M. & Green, H. Stratification and terminal differentiation of cultured epidermal cells. *Nature* **295**, 434–436 (1982).
5. Betschinger, J. & Knoblich, J. A. Dare to be different: asymmetric cell division in *Drosophila*, *C. elegans* and vertebrates. *Curr. Biol.* **14**, R674–R685 (2004).
6. Cowan, C. R. & Hyman, A. A. Asymmetric cell division in *C. elegans*: cortical polarity and spindle positioning. *Annu. Rev. Cell Dev. Biol.* **20**, 427–453 (2004).
7. Roegiers, F. & Jan, Y. N. Asymmetric cell division. *Curr. Opin. Cell Biol.* **16**, 195–205 (2004).
8. Mills, A. A. *et al.* p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398**, 708–713 (1999).
9. Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398**, 714–718 (1999).
10. Mariotti, A. *et al.* EGF-R signalling through Fyn kinase disrupts the function of integrin α 6 β 4 at hemidesmosomes: role in epithelial cell migration and carcinoma invasion. *J. Cell Biol.* **155**, 447–458 (2001).

11. Schlaepfer, D. D. & Hunter, T. Evidence for *in vivo* phosphorylation of the Grb2 SH2-domain binding site on focal adhesion kinase by Src-family protein-tyrosine kinases. *Mol. Cell. Biol.* **16**, 5623–5633 (1996).
12. Tumber, T. *et al.* Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363 (2004).
13. Kraut, R., Chia, W., Jan, L. Y., Jan, Y. N. & Knoblich, J. A. Role of inscuteable in orienting asymmetric cell divisions in *Drosophila*. *Nature* **383**, 50–55 (1996).
14. Parmentier, M. L. *et al.* Rapsynoid/partner of inscuteable controls asymmetric division of larval neuroblasts in *Drosophila*. *J. Neurosci.* **20**, RC84 (2000).
15. Schober, M., Schaefer, M. & Knoblich, J. A. Bazooka recruits Inscuteable to orient asymmetric cell divisions in *Drosophila* neuroblasts. *Nature* **402**, 548–551 (1999).
16. Yu, F., Morin, X., Cai, Y., Yang, X. & Chia, W. Analysis of partner of inscuteable, a novel player of *Drosophila* asymmetric divisions, reveals two distinct steps in inscuteable apical localization. *Cell* **100**, 399–409 (2000).
17. Du, Q. & Macara, I. G. Mammalian Pins is a conformational switch that links NuMA to heterotrimeric G proteins. *Cell* **119**, 503–516 (2004).
18. Du, Q., Stukenberg, P. T. & Macara, I. G. A mammalian Partner of inscuteable binds NuMA and regulates mitotic spindle organization. *Nature Cell Biol.* **3**, 1069–1075 (2001).
19. Gaglio, T., Saredi, A. & Compton, D. A. NuMA is required for the organization of microtubules into aster-like mitotic arrays. *J. Cell Biol.* **131**, 693–708 (1995).
20. Kisurina-Evgenieva, O. *et al.* Multiple mechanisms regulate NuMA dynamics at spindle poles. *J. Cell Sci.* **117**, 6391–6400 (2004).
21. Srinivasan, D. G., Fisk, R. M., Xu, H. & van den Heuvel, S. A complex of Lin-5 and GPR proteins regulates G-protein signalling and spindle function in *C. elegans*. *Genes Dev.* **17**, 1225–1239 (2003).
22. Price, C. M. & Pettijohn, D. E. Redistribution of the nuclear mitotic apparatus protein (NuMA) during mitosis and nuclear assembly. Properties of purified NuMA protein. *Exp. Cell Res.* **166**, 95–317 (1986).
23. Grill, S. W., Howard, J., Schaffer, E., Stelzer, E. H. & Hyman, A. A. The distribution of active force generators controls mitotic spindle position. *Science* **301**, 518–521 (2003).
24. Dowling, J., Yu, Q. C. & Fuchs, E. Beta4 integrin is required for hemidesmosome formation, cell adhesion and cell survival. *J. Cell Biol.* **134**, 559–572 (1996).
25. van der Neut, R., Krimpenfort, P., Calafat, J., Niessen, C. M. & Sonnenberg, A. Epithelial detachment due to absence of hemidesmosomes in integrin $\beta 4$ null mice. *Nature Genet.* **13**, 366–369 (1996).
26. Brakebusch, C. *et al.* Skin and hair follicle integrity is crucially dependent on beta 1 integrin expression on keratinocytes. *EMBO J.* **19**, 3990–4003 (2000).
27. Raghavan, S., Bauer, C., Mundscha, G., Li, Q. & Fuchs, E. Conditional ablation of $\beta 1$ integrin in skin. Severe defects in epidermal proliferation, basement membrane formation, and hair follicle invagination. *J. Cell Biol.* **150**, 1149–1160 (2000).
28. Vasioukhin, V., Bowers, E., Bauer, C., Degenstein, L. & Fuchs, E. Desmoplakin is essential in epidermal sheet formation. *Nature Cell Biol.* **3**, 1076–1085 (2001).
29. Vasioukhin, V., Bauer, C., Degenstein, L., Wise, B. & Fuchs, E. Hyperproliferation and defects in epithelial polarity upon conditional ablation of alpha-catenin in skin. *Cell* **104**, 605–617 (2001).
30. Lu, B., Roegiers, F., Jan, L. Y. & Jan, Y. N. Adherens junctions inhibit asymmetric division in the *Drosophila* epithelium. *Nature* **409**, 522–525 (2001).
31. Kaushik, R., Yu, F., Chia, W., Yang, X. & Bahri, S. Subcellular localization of LGN during mitosis: evidence for its cortical localization in mitotic cell culture systems and its requirement for normal cell cycle progression. *Mol. Biol. Cell* **14**, 3144–3155 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Bornens and A. Mills for reagents; J. Fan for transgenic injections; L. Polak, E. Gonzales and LARC staff for care of the mice; H. Rhee for assistance with flow cytometry; members of the Fuchs laboratory for criticisms; and T. Kapoor for reagents and discussions. E.F. is an Investigator of the Howard Hughes Medical Institute. T.L. is a Jane Coffin Child postdoctoral fellow. This work was supported by a grant from the National Institutes of Health.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to E.F. (fuchslb@rockefeller.edu).

Interaction of phosphorylated c-Jun with TCF4 regulates intestinal cancer development

Abdolrahman S. Nateri¹, Bradley Spencer-Dene^{2,3} & Axel Behrens¹

The proto-oncoprotein c-Jun is a component of the AP-1 transcription factor, the activity of which is augmented in many tumour types¹. An important mechanism in the stimulation of AP-1 function is amino-terminal phosphorylation of c-Jun by the c-Jun N-terminal kinases (JNKs)². Phosphorylated c-Jun is biologically more active, partially because it acquires the ability to interact with binding partners. Here we show that phosphorylated c-Jun interacts with the HMG-box transcription factor TCF4 to form a ternary complex containing c-Jun, TCF4 and β -catenin. Chromatin immunoprecipitation assays revealed JNK-dependent c-Jun–TCF4 interaction on the *c-jun* promoter, and c-Jun and TCF4 cooperatively activated the *c-jun* promoter in reporter assays in a β -catenin-dependent manner. In the *Apc*^{Min} mouse model of intestinal cancer⁶, genetic abrogation of c-Jun N-terminal phosphorylation³ or gut-specific conditional *c-jun* inactivation^{4,5} reduced tumour number and size and prolonged lifespan. Therefore, the phosphorylation-dependent interaction between c-Jun and TCF4 regulates intestinal tumorigenesis by integrating JNK and APC/ β -catenin, two distinct pathways activated by WNT signalling.

Activation of the WNT signalling pathway stimulates growth and mediates developmental signals between cells. In the absence of WNT signal, unstimulated cells regulate β -catenin levels by a multiprotein complex consisting of the adenomatous polyposis coli (APC) tumour suppressor protein, axin and the glycogen synthase kinase GSK3- β , which phosphorylates β -catenin, marking it for subsequent ubiquitination and degradation⁷. Upon docking of the WNT ligand to its frizzled receptor, a cascade of events (termed canonical WNT signalling) destabilizes the degradation complex, allowing unphosphorylated β -catenin levels to accumulate and translocate to the nucleus where β -catenin functions as a cofactor for transcription factors of the T-cell factor/lymphoid enhancing factor (TCF/LEF) family⁷.

Germline mutation of APC in humans results in familial adenomatous polyposis and other tumour syndromes characterized by multiple intestinal adenomas, which invariably transform into cancer⁸. Similarly, mice heterozygous for a nonsense mutation at codon 850 of the *Apc* gene (*Apc*^{Min}) develop multiple intestinal neoplasias due to subsequent loss of the wild-type allele⁶.

WNT signalling can, however, also activate non-canonical, β -catenin-independent pathways. Several pathways of non-canonical WNT signal transduction have been identified, including signalling through calcium flux and activation of the JNKs⁹. c-Jun and AP-1 activity is modified in response to many extracellular stimuli through phosphorylation of c-Jun within its N-terminus by the JNK group of mitogen-activated protein (MAP) kinases². c-Jun N-terminal phosphorylation at serine residues 63 and 73 and threonine residues 91 and 93 within its transactivation domain increases the transcription of target genes, one of which is the *c-jun* gene itself¹⁰.

Mice lacking c-Jun N-terminal phosphorylation generated by a knock-in genetic approach develop normally but show specific defects in stress-induced neuronal apoptosis and oncogenic transformation^{3,11}.

To investigate the mechanism of c-Jun N-terminal phosphorylation-mediated transcriptional regulation, we have previously used a genetic screen in yeast to identify proteins that interact with c-Jun in a phosphorylation-dependent manner (phosphorylation-dependent c-Jun interactors, or PDJs)¹². One of the candidate proteins (PDJ4) that preferentially interacted with N-terminally phosphorylated c-Jun encoded the HMG-box transcription factor TCF4.

Endogenous c-Jun and TCF4 interacted in HCT116 colon cancer cells, and pre-incubation of the polyclonal serum used for immunoprecipitation with recombinant c-Jun prevented interaction, confirming specificity (Fig. 1a). Overexpression experiments in 293T cells confirmed that wild-type c-Jun efficiently bound to TCF4, whereas a mutant c-Jun allele lacking all N-terminal phosphorylation sites (c-Jun^{4A}) interacted less well (Fig. 1b). Further mapping showed that serines 63 and 73, but not threonines 91 and 93, are required for the c-Jun–TCF4 interaction (Supplementary Fig. 1). A TCF4 mutant lacking the N-terminal β -catenin-binding domain retained c-Jun interaction, indicating that the binding sites for β -catenin and c-Jun on TCF4 are distinct (Fig. 1c). β -catenin was detectable in c-Jun immunoprecipitations, but this interaction was significantly reduced by short interfering RNA (siRNA)-mediated knockdown of TCF4, indicating that TCF4 is able to interact simultaneously with c-Jun and β -catenin (Fig. 1d).

To examine the biological function of the c-Jun–TCF4 interaction, the regulation of the *c-jun* promoter was studied. c-Jun autoregulates its transcription via two proximal binding sites (called jun1 and jun2) (Fig. 2a)¹³. In addition, *c-jun* has been identified as one of the most highly induced TCF/ β -catenin target genes in both haematopoietic cells and colon cancer cells^{14,15}; this induction is mediated by a TCF consensus binding site located approximately 3 kilobases (kb) upstream of the jun sites in the human genome (Fig. 2a)¹⁵. Chromatin immunoprecipitation revealed efficient binding of TCF4 to the *c-myc* promoter, a known TCF4 target gene¹⁶, but no binding to control (*Gapdh*, *fra-1* and *Srf*) promoters (Fig. 2b). TCF4 binding to the *c-myc* promoter was insensitive to JNK activation by anisomycin or pharmacological JNK inhibition. TCF4 also bound efficiently to its cognate site in the *c-jun* promoter, but this binding was decreased by JNK inhibition. Notably, the proximal c-Jun binding sites were also detectable in TCF4 chromatin immunoprecipitation in the presence, but not the absence, of JNK activity (Fig. 2b). Moreover, wild-type c-Jun interacted efficiently with the *Cdc2* promoter, an established c-Jun target gene¹⁷, as well as the jun1/2 and the TCF sites of the *c-jun* promoter, but not the TCF site of the *c-myc* promoter (Fig. 2c). Binding of the c-Jun^{4A} mutant protein to the *Cdc2* AP-1 site was similar to wild-type c-Jun, but c-Jun^{4A} binding to both the AP-1 and

¹Mammalian Genetics Laboratory, ²Experimental Pathology Laboratory, CR UK London Research Institute, Lincoln's Inn Fields Laboratories, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. ³Department of Histopathology, Imperial College London, UK.

the TCF sites of the *c-jun* promoter was reduced (Fig. 2c), although c-Jun^{4A} protein was expressed at levels comparable to wild-type c-Jun (Supplementary Fig. 2). β -catenin chromatin immunoprecipitation revealed JNK-independent binding to the TCF sites in the *c-myc* and *c-jun* promoters in HCT116 cells, but not in TIG fibroblasts (Fig. 2d) that lack nuclear β -catenin (Supplementary Fig. 3). JNK inhibition also reduced the interaction of β -catenin with the *c-jun* AP-1 site (Fig. 2d). Therefore c-Jun and TCF4/ β -catenin interact on the *c-jun* promoter *in vivo* in a JNK-dependent manner.

To address the significance of the c-Jun–TCF4 interaction in *c-jun* regulation, a 10-kb murine *c-jun* promoter fragment was fused to a luciferase reporter gene (*c-jun-luc*, Fig. 2a). In NIH3T3 fibroblasts, c-Jun activated *c-jun-luc*, but neither c-Jun^{4A} nor TCF4 did so (Supplementary Fig. 3). JNK inhibition also reduced the interaction of β -catenin with the *c-jun* AP-1 site (Fig. 2d). Therefore c-Jun and TCF4/ β -catenin interact on the *c-jun* promoter *in vivo* in a JNK-dependent manner.

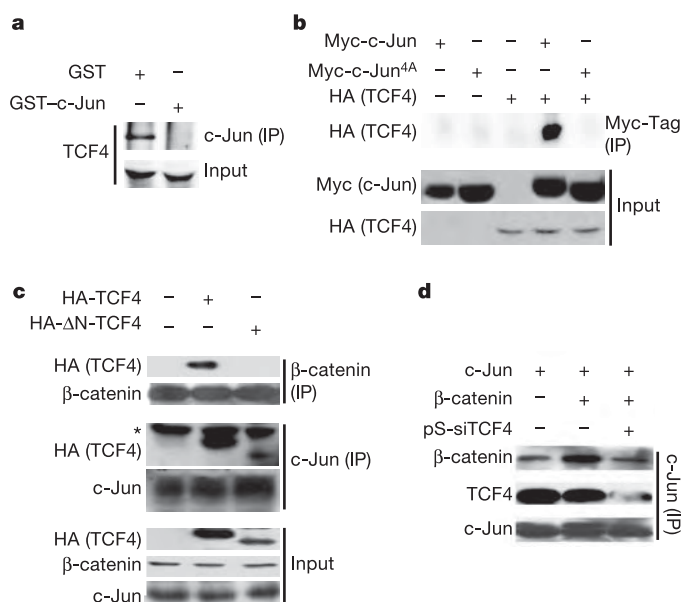


Figure 1 | TCF4 interacts with phosphorylated c-Jun. **a**, TCF4 is present in c-Jun immunoprecipitations (IP). Pre-incubation with GST–c-Jun, but not GST alone, prevents binding. **b**, 293T cells were transfected with Myc-tagged c-Jun, c-Jun^{4A} and haemagglutinin (HA)-TCF4. Myc-Tag immunoprecipitations were performed and TCF4 protein levels were analysed by western blot. **c**, HEK 293T cells were transfected with HA-TCF4 or HA-ΔN-TCF4, c-Jun or β -catenin immunoprecipitations were performed, and TCF4, c-Jun and β -catenin protein levels were analysed by western blotting. An asterisk indicates an unspecific band. **d**, 293T cells were transfected with c-Jun and β -catenin plasmids or a pSUPER vector expressing a short hairpin RNA (shRNA) specific for TCF4 (pS-siTCF4), c-Jun immunoprecipitations were performed, and TCF4, c-Jun and β -catenin protein levels were analysed by western blotting. In all experiments JNK activity was stimulated by treatment of cells with 5 mM anisomycin 1–2 h before immunoprecipitation.

abolish, c-Jun-mediated reporter gene activation (Fig. 2f). Therefore, the TCF and AP-1 sites are both required *in cis* for cooperative transcriptional activation by c-Jun and TCF4/ β -catenin.

To elucidate the physiological relevance of the c-Jun–TCF4 interaction, *Apc*^{Min/+} mice were crossed to *JunAA* mice, in which the two JNK phosphorylation sites on c-Jun required for TCF4 binding—serines 63 and 73 (Supplementary Fig. 1)—are mutated to alanines³. Both control *Apc*^{Min/+}; *c-jun*^{+/+} and *Apc*^{Min/+}; *JunAA*/+ mice succumbed to intestinal cancer at the latest after 19 weeks (average

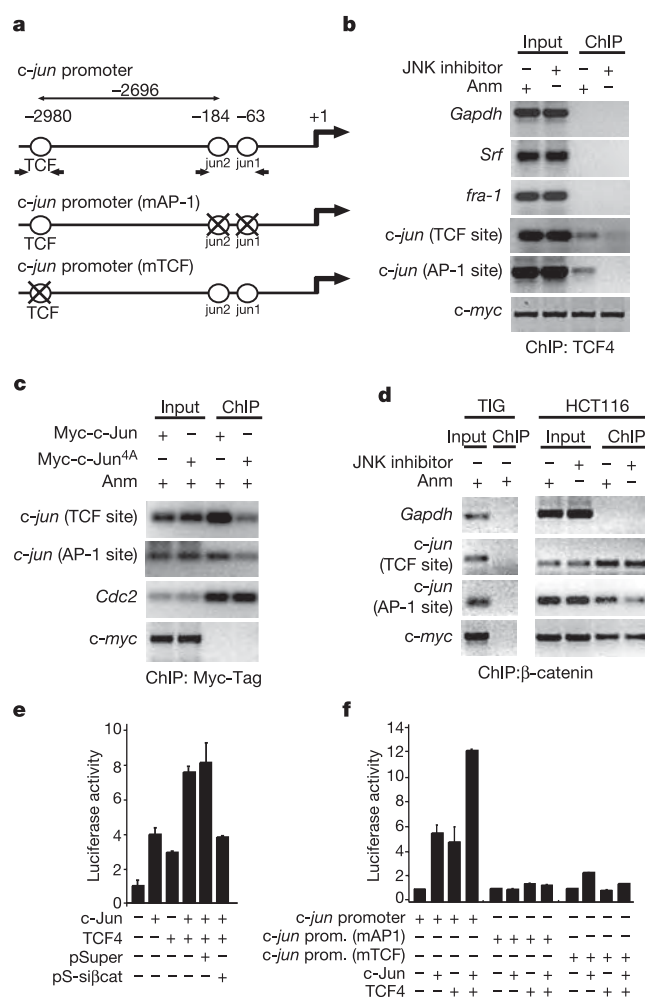


Figure 2 | JNK-dependent interaction between c-Jun and TCF4 regulates the *c-jun* promoter. **a**, Schematic representation of the human *c-jun* promoter. Numbers denote nucleotides relative to the transcriptional start site indicated by a bold arrow (+1). Open circles indicate regulatory DNA elements, crossed circles indicate mutated sites. Arrows indicate approximate positions of PCR primers used in chromatin immunoprecipitation analysis. **b**, TCF4 chromatin immunoprecipitation (ChIP) was performed using HCT116 cells. **c**, Chromatin immunoprecipitation for the Myc tag was performed on HCT116 cells transfected with Myc-tagged c-Jun or c-Jun^{4A}. **d**, β -catenin chromatin immunoprecipitation was performed using TIG or HCT116 cells. Cells were treated for 1 h with JNK inhibitor or anisomycin (Anm) in **b–d**. **e**, HCT116 cells were transfected to express c-Jun, c-Jun^{4A}, TCF4, pSUPER or a pSUPER vector expressing a shRNA specific for β -catenin (pS-si β cat). Activity of a *c-jun* promoter firefly luciferase reporter construct (*c-jun-luc*) relative to a tk-renilla luciferase transfection control (tk-Rluc) is shown. **f**, HCT116 cells were transfected to express c-Jun, c-Jun^{4A}, TCF4, tk-Rluc and either *c-jun-luc*, *c-jun-luc*(mAP1) or *c-jun-luc*(mTCF). The luciferase activity of cells transfected with only *c-jun-luc* and tk-Rluc is arbitrarily set to 1 in **e, f**. Values in **e, f** are mean $\pm$ s.e.m. (standard error of the mean).

15.7 weeks for *c-jun*^{+/+} and 16.1 weeks for *JunAA*/+, but the maximal and average lifespan of *Apc*^{Min/+}; *JunAA*/+ mice was significantly extended (average 23.1 weeks; $P < 0.001$) (Fig. 3a). Quantification of the number of adenomas revealed a moderate reduction throughout the intestine in *Apc*^{Min/+}; *JunAA*/+ mice (Fig. 3b, c, f; see also Supplementary Fig. 6). In *Apc*^{Min/+}; *JunAA*/+ mice most adenomas were small (Fig. 3d, e, g), and the average size of *Apc*^{Min/+}; *JunAA*/+ adenomas was significantly reduced (Fig. 3h; $P < 0.01$). Therefore, the *JunAA* mutation greatly reduces the tumour burden of *Apc*^{Min/+} mice.

Whereas the amount of cell death was comparable between *Apc*^{Min/+}; *c-jun*^{+/+} and *Apc*^{Min/+}; *JunAA*/+ adenomas (data not shown), all c-Jun-positive cells were proliferating as judged by 5-bromodeoxyuridine (BrdU) incorporation (Supplementary Fig. 7). *Apc*^{Min/+}; *JunAA*/+ adenoma cells had a decreased proliferation index (Fig. 3i), indicating that reduced proliferation is a main reason for the decrease in tumour size. In agreement with the *c-jun* promoter analysis (Fig. 2), *c-jun* messenger RNA and c-Jun protein levels were reduced in adenomas lacking c-Jun N-terminal phosphorylation, but the expression of the TCF4 targets c-Myc and cyclinD1 (refs 16, 18), and total or phosphorylated JNK protein levels

were unaffected (Fig. 3j, k).

To test directly the importance of *c-jun* as a target of the c-Jun–TCF4 interaction during intestinal cancer development, a floxed allele of *c-jun* (*c-jun*^{fl}) was inactivated specifically in the gut using *villin-cre* transgenic mice (*c-jun*^{ΔG} mice; Fig. 4a–c). The absence of *c-jun* had a more pronounced effect on the lifespan of *Apc*^{Min/+} mice than the *JunAA* mutation, because *Apc*^{Min/+}; *c-jun*^{ΔG} mice did not display any clinical sign of cancer development at 9 months of age ($n = 8$). At 18 weeks, *Apc*^{Min/+}; *c-jun*^{fl/fl} mice were at the end stage of disease and displayed numerous large adenomas (Fig. 4d). In contrast, *Apc*^{Min/+}; *c-jun*^{ΔG} mice were characterized by the presence of cystic structures (Fig. 4e). c-Jun could not be detected in the gut epithelium of *Apc*^{Min/+}; *c-jun*^{ΔG} mice, but was highly expressed in *Apc*^{Min/+}; *c-jun*^{fl/fl} adenomas (Fig. 4f, g). *Apc*^{Min/+}; *c-jun*^{fl/fl} tumours showed accumulation of β-catenin in the nucleus, as did the cysts present in *Apc*^{Min/+}; *c-jun*^{ΔG} mice (Fig. 4h, i; arrowheads). These cysts are not present in wild-type or *c-jun*^{ΔG} mice, suggesting that they are caused by loss of Apc function and that β-catenin signalling is unable to trigger tumour development in the absence of c-Jun. Expression of CD44, which is a c-Jun target gene¹⁹ and induced in intestinal tumours^{20,21}, was also greatly reduced in adenomas of *Apc*^{Min/+}; *c-jun*^{ΔG} mice (Fig. 4c, j, k). Thus, whereas the absence of

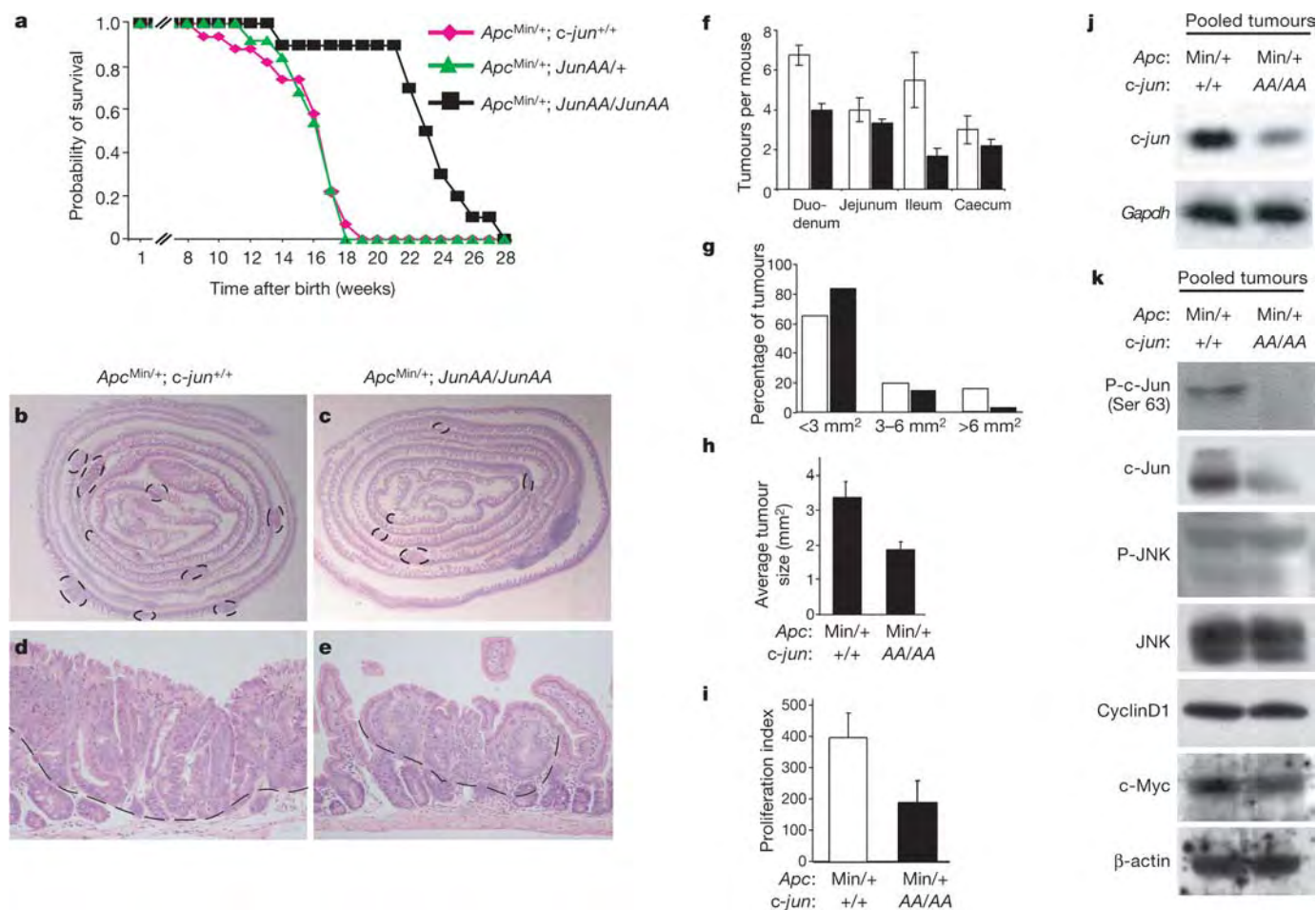


Figure 3 | Absence of c-Jun N-terminal phosphorylation attenuates intestinal cancer development in *Apc*^{Min} mice. **a**, Kaplan–Meier survival plots of *Apc*^{Min/+}; *c-jun*^{+/+} ($n = 16$), *Apc*^{Min/+}; *JunAA*/+ ($n = 13$) and *Apc*^{Min/+}; *JunAA*/+ ($n = 10$) mice. The x axis indicates the time (in weeks) until mice had to be killed due to severe weight loss and anaemia. **b, c**, Pictures of whole intestines. Dashed circles indicate adenomas. **d, e**, Haematoxylin and eosin (H&E) staining of intestinal sections. Dashed lines indicate intestinal adenomas. **f**, Average tumour number grouped

according to intestinal location in *Apc*^{Min/+}; *c-jun*^{+/+} ($n = 100$, white bars) and *Apc*^{Min/+}; *JunAA*/+ mice ($n = 49$, black bars) at the age of 11 weeks. **g**, Quantification (in per cent) of tumour size distribution. White columns, *Apc*^{Min/+}; *c-jun*^{+/+} mice; black columns, *Apc*^{Min/+}; *JunAA*/+ mice. **h**, Average tumour size at the age of 11 weeks; $P < 0.01$. **i**, Quantification of the proliferation index. **j**, Northern blot analysis of *c-jun* and *Gapdh* expression (loading control). **k**, Western blot analysis of indicated proteins. Values in **f**, **h**, **i** are mean $\pm$ s.e.m.

c-Jun N-terminal phosphorylation delayed intestinal tumorigenesis, inactivation of *c-jun* seems to protect *Apc*^{Min/+} mice from intestinal cancer.

The phosphorylation-dependent c-Jun–TCF4 interaction is a molecular mechanism to integrate signals that cause activation of the TCF/β-catenin pathway and JNK. The transcriptional

cooperation between TCF4 and c-Jun is mediated by activated β-catenin and is dependent on the presence of both the TCF and AP-1 *c-jun* promoter elements (Fig. 2), suggesting that the phospho-c-Jun–TCF4 interaction may stimulate transcription by recruiting β-catenin to the proximal AP-1 elements close to the transcriptional start site (Fig. 4l).

The normal development and lifespan of *JunAA* homozygous mice suggests that inhibition of the phospho-c-Jun–TCF4 interaction may be a promising therapeutic strategy to treat human intestinal cancer. Moreover, small molecule and peptide-based JNK inhibitors have been identified^{22,23}, and it will be interesting to determine whether these reagents have a beneficial therapeutic effect on colon cancer progression.

METHODS

Mouse lines. *JunAA/JunAA* (ref. 3), *c-jun*^{fl} (ref. 4), *villin-cre* transgenic mice (ref. 5) and *Apc*^{Min} mice⁶ were described previously. *Apc*^{Min} mice were on a C57/BL6 genetic background. *JunAA/JunAA* and *c-jun*^{fl} mice were backcrossed at least five times, and *villin-cre* transgenic mice at least three times, to a C57/BL6 background. In all experiments only littermates were used and compared.

Biochemical analysis. The siRNA oligonucleotides targeting TCF4 and β-catenin were from the NK1 RNAi library cloned in pRetroSUPER. DNA transfections, immunoprecipitations, northern and western analysis and luciferase reporter gene assays were performed as described previously¹². Details are described in the Supplementary Information.

Chromatin immunoprecipitation. Chromatin immunoprecipitation analysis was performed as described previously²⁴. In some cases, cells were treated 1 h before collection with 25 μM JNK Inhibitor II (Calbiochem) and 5 μM anisomycin (Sigma). Immunoprecipitations were carried out with conjugated agarose beads for anti-MYC (Sigma) and anti-TCF4 (Santa Cruz) antibodies. The oligonucleotide sequences used to amplify the DNA fragments are available upon request.

Histological analysis. Tumour number and size were determined by image analysis of haematoxylin and eosin-stained tissue sections. Monoclonal anti-BrdU-fluorescein isothiocyanate conjugate (Becton Dickinson) and antibodies specific for TCF-4 (Upstate), β-catenin (Cell Signaling), CD44 (Chemicon) and c-Jun (both Santa Cruz) were used for immunohistochemistry and immunofluorescence. Technical details are described in the Supplementary Information.

Received 10 March; accepted 14 June 2005.

Published online 10 July 2005.

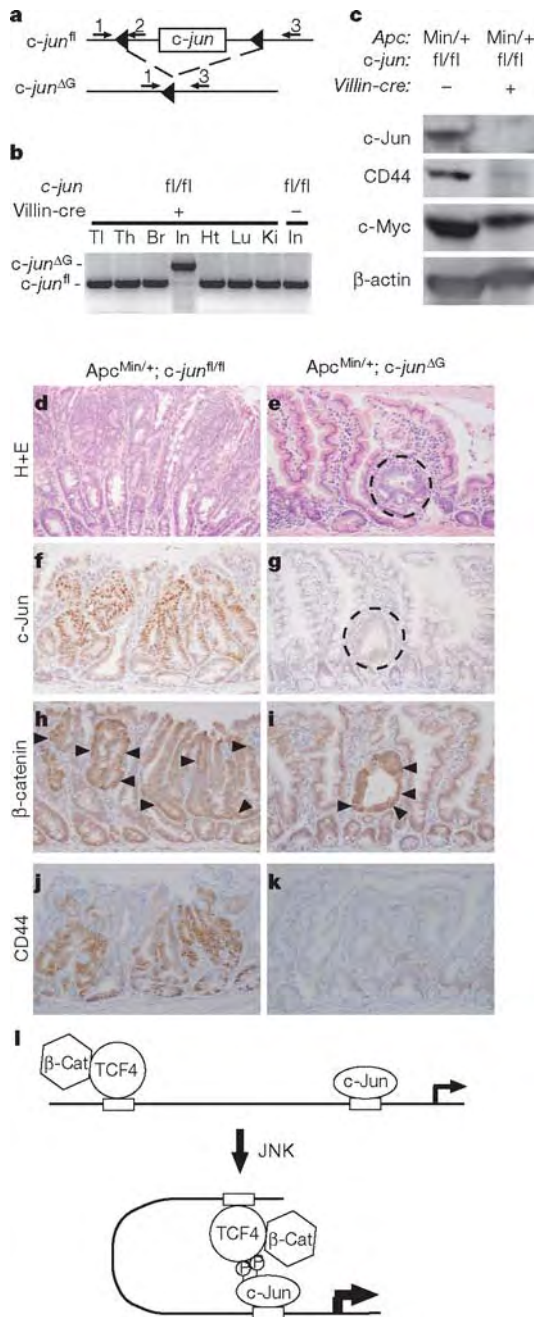


Figure 4 | c-Jun is required for intestinal cancer development in *Apc*^{Min/+} mice. **a**, Scheme of the floxed *c-jun* allele before and after *cre* recombination. Arrows indicate approximate position of PCR primers used. **b**, PCR analysis of genomic DNA. Br, brain; Ht, heart; In, intestine; Ki, Kidney; Lu, lung; Th, thymus; Tl, tail. **c**, Western blot analysis of *Apc*^{Min/+} *c-jun*^{fl/fl} control and *Apc*^{Min/+} *c-jun*^{ΔG} intestines. **d, e**, Haematoxylin and eosin (H+E) staining of intestinal sections (18-week-old mice). The dashed circle in **e** indicates a cyst. **f–k**, Immunohistochemistry for c-Jun (**f, g**), β-catenin (**h, i**) and CD44 (**j, k**) on intestinal sections. The dashed circle in **g** indicates a cyst; arrowheads in **h** and **i** indicate cells with nuclear β-catenin staining. **k**, Schematic representation of a JNK-dependent interaction between c-Jun and TCF4/β-catenin on the *c-jun* promoter.

1. Eferl, R. & Wagner, E. F. AP-1: a double-edged sword in tumorigenesis. *Nature Rev. Cancer* **3**, 859–868 (2003).
2. Davis, R. J. Signal transduction by the JNK group of MAP kinases. *Cell* **103**, 239–252 (2000).
3. Behrens, A., Sibilio, M. & Wagner, E. F. Amino-terminal phosphorylation of c-Jun regulates stress-induced apoptosis and cellular proliferation. *Nature Genet.* **21**, 326–329 (1999).
4. Behrens, A. et al. Impaired postnatal hepatocyte proliferation and liver regeneration in mice lacking c-jun in the liver. *EMBO J.* **21**, 1782–1790 (2002).
5. el Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).
6. Moser, A. R. et al. *Apc*^{Min}, a mutation in the murine *Apc* gene, predisposes to mammary carcinomas and focal alveolar hyperplasias. *Proc. Natl Acad. Sci. USA* **90**, 8977–8981 (1993).
7. Giles, R. H., van Es, J. H. & Clevers, H. Caught up in a Wnt storm: Wnt signalling in cancer. *Biochim. Biophys. Acta* **1653**, 1–24 (2003).
8. Fodde, R. The APC gene in colorectal cancer. *Eur. J. Cancer* **38**, 867–871 (2002).
9. Veeman, M. T., Axelrod, J. D. & Moon, R. T. A second canon. Functions and mechanisms of β-catenin-independent Wnt signalling. *Dev. Cell* **5**, 367–377 (2003).
10. Angel, P., Hattori, K., Smeal, T. & Karin, M. The *jun* proto-oncogene is positively autoregulated by its product, Jun/AP1. *Cell* **55**, 875–885 (1988).
11. Behrens, A., Jochum, W., Sibilio, M. & Wagner, E. F. Oncogenic transformation by ras and fos is mediated by c-Jun N-terminal phosphorylation. *Oncogene* **19**, 2657–2663 (2000).
12. Nateri, A. S., Riera-Sans, L., Da Costa, C. & Behrens, A. The ubiquitin ligase SCFFbw7 antagonizes apoptotic JNK signalling. *Science* **303**, 1374–1378 (2004).
13. Stein, B. et al. Ultraviolet-radiation induced c-jun gene transcription: two AP-1 like binding sites mediate the response. *Photochem. Photobiol.* **55**, 409–415 (1992).
14. Staal, F. J. et al. Wnt target genes identified by DNA microarrays in immature CD34+ thymocytes regulate proliferation and cell adhesion. *J. Immunol.* **172**, 1099–1108 (2004).

15. Mann, B. *et al.* Target genes of β -catenin-T cell-factor/lymphoid-enhancer-factor signalling in human colorectal carcinomas. *Proc. Natl Acad. Sci. USA* **96**, 1603–1608 (1999).
16. He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512 (1998).
17. Wada, T. *et al.* MKK7 couples stress signalling to G2/M cell-cycle progression and cellular senescence. *Nature Cell Biol.* **6**, 215–226 (2004).
18. Tetsu, O. & McCormick, F. β -catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
19. Raivich, G. *et al.* The AP-1 transcription factor c-Jun is required for efficient axonal regeneration. *Neuron* **43**, 57–67 (2004).
20. Wielenga, V. J. *et al.* Expression of CD44 in Apc and Tcf mutant mice implies regulation by the WNT pathway. *Am. J. Pathol.* **154**, 515–523 (1999).
21. Sansom, O. J. *et al.* Loss of Apc *in vivo* immediately perturbs Wnt signalling, differentiation, and migration. *Genes Dev.* **18**, 1385–1390 (2004).
22. Bogoyevitch, M. A., Boehm, I., Oakley, A., Ketterman, A. J. & Barr, R. K. Targeting the JNK MAPK cascade for inhibition: basic science and therapeutic potential. *Biochim. Biophys. Acta* **1697**, 89–101 (2004).
23. Borsello, T. *et al.* A peptide inhibitor of c-Jun N-terminal kinase protects

against excitotoxicity and cerebral ischemia. *Nature Med* **9**, 1180–1186 (2003).

24. Alberts, A. S., Geneste, O. & Treisman, R. Activation of SRF-regulated chromosomal templates by Rho-family GTPases requires a signal that also induces H4 hyperacetylation. *Cell* **92**, 475–487 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank T. Akiyama, K. Niemann, G. Peters and F. Watt for providing essential reagents. We are grateful to A. Gilkar, W. Jochum, E. Nye and G. Stamp for technical help and advice on histology. We thank K. Sabapathy, I. Tomlinson and R. Treisman for critical reading of the manuscript. The London Research Institute is funded by CR UK.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare competing financial interests: details accompany the paper on www.nature.com/nature. Correspondence and requests for materials should be addressed to A.B. (axel.behrens@cancer.org.uk).

LETTERS

A mechanistic principle for proton pumping by cytochrome *c* oxidase

Kristina Faxén¹, Gwen Gilderson¹, Pia Ädelroth¹ & Peter Brzezinski¹

In aerobic organisms, cellular respiration involves electron transfer to oxygen through a series of membrane-bound protein complexes. The process maintains a transmembrane electrochemical proton gradient that is used, for example, in the synthesis of ATP. In mitochondria and many bacteria, the last enzyme complex in the electron transfer chain is cytochrome *c* oxidase (Cyt*c*O), which catalyses the four-electron reduction of O₂ to H₂O using electrons delivered by a water-soluble donor, cytochrome *c*^{1–7}. The electron transfer through Cyt*c*O, accompanied by proton uptake to form H₂O drives the physical movement (pumping) of four protons across the membrane⁸ per reduced O₂. So far, the molecular mechanism of such proton pumping driven by electron transfer has not been determined in any biological system. Here we show that proton pumping in Cyt*c*O is mechanistically coupled to proton transfer to O₂ at the catalytic site, rather than to internal electron transfer. This scenario suggests a principle by which redox-driven proton pumps might operate and puts considerable constraints on possible molecular mechanisms by which Cyt*c*O translocates protons.

During catalytic turnover, cytochrome *c* docks to Cyt*c*O at the electrically positive (P-) side of the membrane and donates one electron at a time to the primary electron acceptor, copper A (Cu_A). The electron is then transferred consecutively to a haem group (haem *a*) and to the catalytic site, which consists of another haem group (haem *a*₃) and a copper ion (Cu_B) (Fig. 1a). Because cytochrome *c* is a one-electron donor and four electrons are required to reduce O₂ to H₂O, during catalytic turnover Cyt*c*O cycles between distinct states with an increasing number of electrons transferred to the catalytic site (haem *a*₃–Cu_B). These states are named by one-letter codes where the superscript indicates the number of electrons transferred to the catalytic site. The consecutive addition of two electrons to the oxidised Cyt*c*O (state O⁰) results first in reduction of Cu_B (state E¹) and then also haem *a*₃ (Fig. 1b, state R²). In state R² haem *a*₃ binds O₂, which is reduced to form a state that is called 'peroxy' (P², also called P_M). There is consensus in the literature that on binding of O₂ to the reduced catalytic site (R² → P² transition) the O–O bond is broken, which activates the oxygen molecule⁹. Addition of the third and fourth electrons results in consecutive formation of the 'ferry' (F³) and oxidised (O⁴⁽⁰⁾) states, which completes the cycle (note that states O⁰ and O⁴ are equivalent). Each of the electron transfers during the O⁰⁽⁴⁾ → E¹, E¹ → R², P² → F³ and F³ → O⁴⁽⁰⁾ transitions is coupled to proton uptake from the membrane negative (N-) side (Fig. 1b) and has been suggested to be associated with proton pumping² (Fig. 1a).

One of the central questions when addressing the mechanism of energy conservation in the respiratory chain is how the O₂-reduction reaction in Cyt*c*O is energetically and mechanistically linked to proton pumping. In mechanistically simpler systems, such as bacteriorhodopsin, each proton-pumping event is driven by a structural

change that occurs on absorption of a photon by a chromophore¹⁰. In contrast, in Cyt*c*O the driving reaction involves a multi-step chemical reaction as well as electron and proton transfer to the catalytic site, which complicates studies of the mechanism.

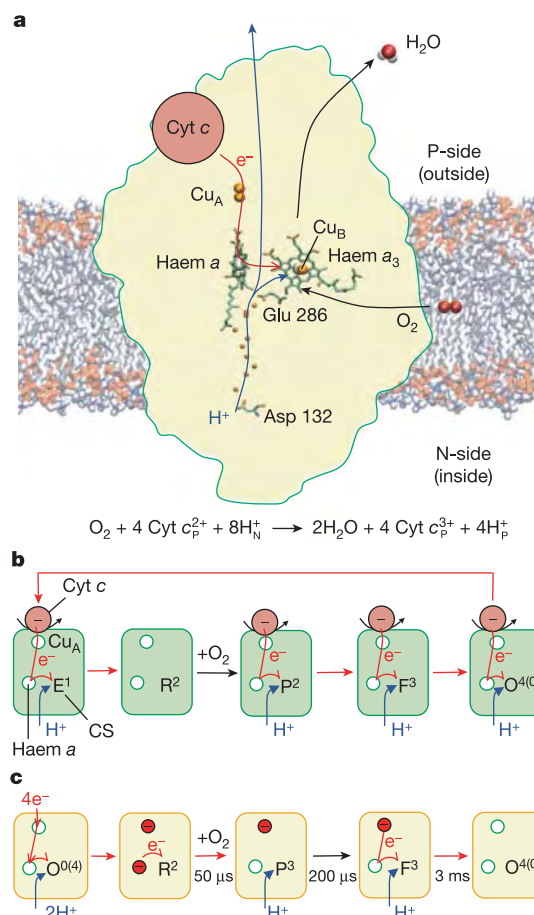


Figure 1 | Structural and functional properties of cytochrome *c* oxidase.

a, The redox-active cofactors, the D-proton-transfer pathway (from Asp 132 to Glu 286) and the reaction catalysed by Cyt*c*O (PDB number 1M56, ref. 24). N and P refer to the two sides of the membrane. **b**, The catalytic cycle of Cyt*c*O. **c**, The reaction of the four-electron reduced Cyt*c*O with O₂ (studied in this work). In **b** and **c** the different states of Cyt*c*O are indicated by one-letter codes where the superscript indicates the number of electrons transferred to the catalytic site (CS). The red and blue arrows indicate electron and proton-transfer reactions, respectively. Cyt *c* indicates the electron donor cytochrome *c*.

¹Department of Biochemistry and Biophysics, The Arrhenius Laboratories for Natural Sciences, Stockholm University, SE-106 91 Stockholm, Sweden.

To study proton pumping in CytcO, we have used the 'flow-flash' technique. The vesicle-reconstituted enzyme is first reduced by four electrons (that is, each of the redox sites is reduced), and carbon monoxide is bound to haem a_3 . The CytcO–CO complex is mixed with an O_2 -containing solution and the reaction with O_2 is initiated by flash-induced dissociation of the CO ligand. On binding of O_2 to haem a_3 , electron transfer from the nearby haem a to the catalytic site ($\tau \approx 50 \mu s$) occurs simultaneously with breaking of the O–O bond (Fig. 1c). As in the case described above, the state that is formed at the catalytic site is also called peroxy (denoted P^3), but because the electron transfer is not charge compensated by proton uptake^{11,12} it carries one extra negative charge in excess of P^2 . In the subsequent reaction step, the excess negative charge in P^3 is neutralized by proton uptake to the catalytic site forming the ferryl state (F^3), with a time constant of $\sim 200 \mu s$. Thus, when the reaction of CytcO with O_2 is initiated in the four-electron reduced enzyme, the electron transfer precedes the proton transfer and the $P^3 \rightarrow F^3$ transition is only associated with proton uptake (compare Fig. 1b with 1c). In the

last step, which is equivalent to the corresponding reaction during turnover, an electron and a proton are transferred to the catalytic site forming state $O^{4(0)}$ ($\tau \approx 3 ms$).

The results from previous studies showed that protons are released to the outside of liposomes containing mitochondrial CytcO concomitantly with the $F^3 \rightarrow O^{4(0)}$ transition¹³, however, no release was detected on the time scale of $\sim 200 \mu s$, that is during the $P^3 \rightarrow F^3$ transition. In contrast, both transitions have been found to be associated with movement of charge across the CytcO, in part attributed to pumping of one proton in each transition^{14,15}. This apparent discrepancy may be rationalized if protons are pumped by CytcO on the $200 \mu s$ time scale, but are then retained at the surface before being released to solution. Alternatively, it has been speculated that protons are not pumped during the $P^3 \rightarrow F^3$ transition, which would mean that the origin of the observed voltage changes must be re-interpreted (for example see ref. 1). The issue is central to the understanding of the proton-pumping mechanism of CytcO, because the $P^3 \rightarrow F^3$ transition is unique in that it is not associated with any simultaneous electron transfer to the catalytic site (Fig. 1c).

In this study we have followed (in real time) proton release to the outside or uptake from the inside of small unilamellar vesicles (SUVs) when the four-electron reduced CytcO (cytochrome aa_3) from *Rhodospirillum rubrum* reacts with O_2 . Figure 2a shows absorbance changes of the dye phenol red added to the outside of the CytcO-SUVs. As seen in the figure there was a biphasic decrease in the dye absorbance with the same time constants as those of the $P^3 \rightarrow F^3$ ($\sim 200 \mu s$) and $F^3 \rightarrow O^{4(0)}$ transitions ($\sim 3 ms$), corresponding to the release of approximately one proton in each transition. The absorbance change of the dye enclosed within the CytcO-SUVs is shown in Fig. 2b. A biphasic increase in absorbance with time constants of $\sim 200 \mu s$ and $\sim 3 ms$ with approximately equal amplitudes was observed, attributed to proton uptake from the inside medium

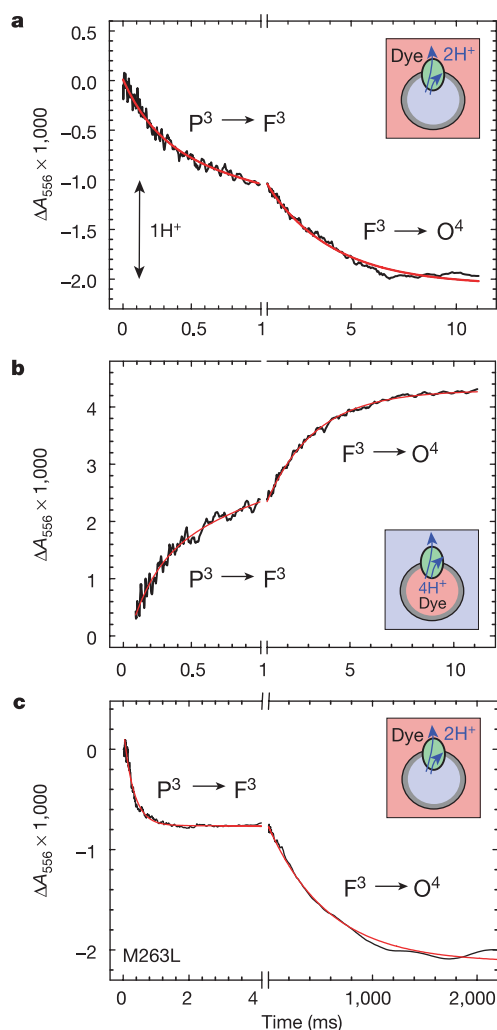


Figure 2 | Absorbance changes of the dye phenol red associated with proton release and uptake on reaction of the four-electron reduced CytcO with O_2 . **a**, Proton release to the P-side solution. The trace shown is the difference between traces obtained without and with buffer. The absorbance amplitude corresponding to the release of $1H^+$ (arrow) is indicated. **b**, Proton uptake from the N-side solution. **c**, Proton release by the M263L CytcO mutant to the P-side solution. The solid lines are fits of the data with a sum of two exponential functions with time constant of $\sim 200 \mu s$ and $\sim 3 ms$ (**a** and **b**) or $\sim 300 \mu s$ and $\sim 600 ms$ (**c**). For experimental conditions see the Methods section.

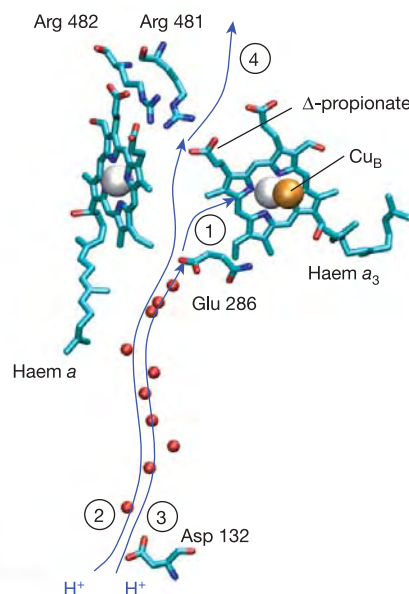


Figure 3 | A schematic mechanism that accounts for proton pumping without simultaneous electron transfer. Electron transfer results in formation of a proton acceptor having a high pK_a at the catalytic site. A proton is transferred to this acceptor from Glu 286 (1), which on deprotonation changes configuration, linked to a structural change that leads to an increase in the pK_a of a protonatable group around the Δ -propionate of haem a_3 and Arg 481/482. This proton acceptor, as well as Glu 286, is rapidly protonated from the N-side (2 and 3). The Glu 286 returns to its original configuration, which is also associated with a structural relaxation of the residues around the Δ -propionate, resulting in proton release to the P-side (4). The schematic diagram is based on ref. 25.

during the $P^3 \rightarrow F^3$ and $F^3 \rightarrow O^{4(0)}$ transitions, respectively. Because it has been shown that each of these transitions is associated with a net uptake of one proton (as measured in detergent solution^{11,12}), two protons are taken up from the solution inside the Cyt_cO-SUVs and one proton is released to the outside bulk solution. A comparison of the data in Fig. 2a and 2b shows that there is no delay between proton uptake from the inside and release to the outside of the Cyt_cO-SUVs.

Although the $P^3 \rightarrow F^3$ transition is not associated with electron transfer to the catalytic site, it occurs on the same time scale as a fractional electron transfer from Cu_A to haem *a* (ref. 12; not shown in Fig. 1c). To investigate whether this electron transfer is important for proton pumping, we examined proton release during the O₂-reduction reaction in a structural variant of Cyt_cO (M263L Cyt_cO; ref. 16), in which the $P^3 \rightarrow F^3$ transition displays about the same time constant as in the native Cyt_cO but Cu_A remains reduced on the time scale of the transition¹⁷. In addition, the $F^3 \rightarrow O^{4(0)}$ transition is slowed from ~3 ms to ~600 ms (ref. 17) in M263L Cyt_cO, which results in a clearer temporal separation of the $P^3 \rightarrow F^3$ and $F^3 \rightarrow O^{4(0)}$ transitions. In M263L Cyt_cO, the $P^3 \rightarrow F^3$ transition was associated with proton release to the outside bulk solution with a time constant of ~300 μs (Fig. 2c). Thus, this proton release is not coupled in time to any electron movement within the Cyt_cO. A second proton was released with a time constant of ~600 ms, concomitantly with the oxidation of Cu_A and formation of O⁴⁽⁰⁾.

The reason why the 200-μs proton release was not observed in previous studies with the mitochondrial Cyt_cO (ref. 13) is not known. One possible explanation is that the mitochondrial Cyt_cO is larger than that from *R. sphaeroides* and therefore has a much larger number of buffering surface groups that may delay the proton release.

As outlined above, electrons enter one at a time through Cu_A and are transferred via haem *a* to the catalytic site during Cyt_cO turnover. Accordingly, a common theme of many proposed mechanisms for proton pumping has been the consecutive reduction and oxidation of a redox site (for example haem *a*), linked to the uptake, internal transfer and release of protons^{18–22}. However, as noted above, one problem associated with such mechanisms is proton pumping during the $P^3 \rightarrow F^3$ transition ($\tau \approx 200 \mu s$), which takes place in the absence of any internal electron transfer. Even if we assume that this transition is unique for the reaction of the four-electron reduced Cyt_cO with O₂, it is reasonable to expect that the pumping mechanism is the same in every step of the reaction cycle. Therefore, when discussing a molecular mechanism of proton pumping we must consider a scenario where protons are pumped without simultaneous internal electron transfer. Also, any model for proton pumping in Cyt_cO needs to take into account that during the reaction of Cyt_cO with O₂ the protons used for O₂ reduction (substrate protons) as well as the protons to be pumped are transferred along the same specific trajectory (called the D pathway), starting near Asp 132 and ending near Glu 286 (Fig. 1a; reviewed in ref. 1). Thus, two protons are transferred through the D pathway for each transition associated with proton pumping. The rate-limiting step for this process was found to be the proton transfer from Glu 286 to the catalytic site^{6,23}, where the deprotonation of Glu 286 presumably leads to local structural changes^{24,25}. These observations were summarized in the form of a molecular mechanism outlined in Fig. 3 (refs 23, 25). According to this scenario, in every reaction step associated with proton pumping, a proton is initially transferred from Glu 286 to the catalytic site, which results in a local structural change that triggers proton translocation across the membrane. This model is further supported by the findings presented here, because it does not involve a direct link of proton pumping to electron transfer (see also refs 26, 27).

In conclusion, we have shown that in one specific reaction step of the Cyt_cO catalytic cycle proton uptake from the N-side and release to the P-side of the membrane (proton pumping) is not linked to simultaneous electron transfer. Instead, proton pumping occurs

simultaneously with proton uptake to the catalytic site during the chemical reaction, which suggests a principle for how this redox-driven proton pump works.

METHODS

Reconstitution of Cyt_cO into liposomes. For reconstitution of Cyt_cO in small unilamellar vesicles (SUVs, 300–400 Å diameter) Cyt_cO was diluted to 4–8 μM in the solution desired inside the vesicles (see below) and 4% (w/v) cholate. A solution of 1-α-phosphatidylcholine (type II-S from soybean, Sigma) was prepared at a concentration of 40 mg ml⁻¹ with 2% (w/v) cholate in the solution desired inside the vesicles. After sonication the liposomes were mixed with the Cyt_cO sample at a ratio of 1:1, and detergent was removed using Bio-Beads (Bio-Rad Laboratories) and a PD-10 column (Amersham Biosciences). The pH was adjusted to a desired value.

The proton permeability of the SUVs was determined by measuring the respiratory-control ratio, that is the Cyt_cO activity in the presence and absence of the ionophores valinomycin (5 μM) and FCCP (carbonyl cyanide-p-(trifluoromethoxy)phenylhydrazone, 10 μM). The respiratory-control ratio was typically around 10, indicating sufficiently tight vesicles.

Proton pumping measurements. The Cyt_cO-SUVs were transferred to an anaerobic cuvette, air was removed and replaced by nitrogen using a vacuum line and the Cyt_cO was reduced by addition of ascorbate and hexaammineruthenium(II)chloride (Acros). Finally, nitrogen was removed and replaced by carbon monoxide. This treatment resulted in four-electron reduced Cyt_cO with CO bound to haem *a*₃. The sample was transferred anaerobically to one of the drive syringes of a flow-flash apparatus (Applied Photophysics)²⁸. The other syringe was loaded with an oxygen-saturated solution. The solutions were mixed at a ratio of 1:1 and about 300 ms after mixing the reaction was initiated simultaneously in the whole population by photo-dissociation of CO by a laser flash (~7 ns, 100 mJ, Quantel, Brilliant B), allowing oxygen to bind to haem *a*₃. Proton uptake and release were monitored using the pH-sensitive dye phenol red, by measuring transient absorbance changes at 556 nm. The dye was either added to the oxygen-saturated solution, or when present inside the SUVs it was added to the lipid and Cyt_cO solutions before sonication.

The number of protons released to the outside of the SUVs on reaction of the reduced Cyt_cO with O₂ was determined from a ratio of the signals obtained without and with the ionophore FCCP. Because in the latter case the SUVs were permeable to protons, the net uptake of approximately two protons¹² was observed on reaction of the reduced Cyt_cO with O₂ (in this case there was no signal from protons that were pumped). The experimental conditions were (after mixing): 1–2 μM Cyt_cO, 50 μM EDTA, 100 mM KCl, 1 mM ascorbate, 0.5 μM hexaammineruthenium(II)chloride, 0.6 mM O₂ and 0.5 mM CO at 22 °C. When measuring proton release to the outside of the Cyt_cO-SUVs the solution also contained: 25 mM sucrose and 62 μM phenol red at pH 7.6 (the solution inside the SUVs contained: 100 mM KCl and 25 mM HEPES at pH 7.6). When measuring proton uptake from the inside of the SUVs the solution also contained: 100 mM KCl, 20 mM sucrose, 50 μM EDTA and 5 mM HEPES at pH 8.2 (the solution inside the SUVs contained: 100 mM KCl and 25 mM phenol red at pH 8.2). The pH was higher when measuring proton uptake from the inside of the SUVs than when measuring proton release to the outside because the pK_a of the dye changed from 7.5 to 8.8 when enclosed within the SUVs. Typically, an average of 20–30 traces was collected. The cuvette path length was 1 cm. Figures 1 and 3 were prepared using the VMD software²⁹ and the membrane in Fig. 1 is from ref. 30.

Received 16 February; accepted 9 June 2005.

- Gennis, R. B. Coupled proton and electron transfer reactions in cytochrome oxidase. *Front. Biosci.* **9**, 581–591 (2004).
- Wikström, M. Cytochrome *c* oxidase: 25 years of the elusive proton pump. *Biochim. Biophys. Acta* **1655**, 241–247 (2004).
- Mills, D. A. & Ferguson-Miller, S. Understanding the mechanism of proton movement linked to oxygen reduction in cytochrome *c* oxidase: lessons from other proteins. *FEBS Lett.* **545**, 47–51 (2003).
- Michel, H., Behr, J., Harrenga, A. & Kannt, A. Cytochrome *c* oxidase: structure and spectroscopy. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 329–356 (1998).
- Rich, P. R. The molecular machinery of Keilin's respiratory chain. *Biochem. Soc. Trans.* **31**, 1095–1105 (2003).
- Brzezinski, P. Redox-driven membrane-bound proton pumps. *Trends Biochem. Sci.* **29**, 380–387 (2004).
- Brunori, M., Giuffrè, A. & Sarti, P. Cytochrome *c* oxidase, ligands and electrons. *J. Inorg. Biochem.* **99**, 324–336 (2005).
- Wikström, M. K. F. Proton pump coupled to cytochrome *c* oxidase in mitochondria. *Nature* **266**, 271–273 (1977).
- Babcock, G. T. How oxygen is activated and reduced in respiration. *Proc. Natl Acad. Sci. USA* **96**, 12971–12973 (1999).

10. Lanyi, J. K. & Luecke, H. Bacteriorhodopsin. *Curr. Opin. Struct. Biol.* **11**, 415–419 (2001).
11. Paula, S., Sucheta, A., Szundi, I. & Einarsdóttir, Ó. Proton and electron transfer during the reduction of molecular oxygen by fully reduced cytochrome c oxidase: a flow-flash investigation using optical multichannel detection. *Biochemistry* **38**, 3025–3033 (1999).
12. Ådelroth, P., Ek, M. & Brzezinski, P. Factors determining electron-transfer rates in cytochrome c oxidase: investigation of the oxygen reaction in the *R. sphaeroides* and bovine enzymes. *Biochim. Biophys. Acta* **1367**, 107–117 (1998).
13. Oliveberg, M., Hallén, S. & Nilsson, T. Uptake and release of protons during the reaction between cytochrome c oxidase and molecular oxygen: a flow-flash investigation. *Biochemistry* **30**, 436–440 (1991).
14. Zaslavsky, D., Kaulen, A. D., Smirnova, I. A., Vygodina, T. & Konstantinov, A. A. Flash-induced membrane potential generation by cytochrome c oxidase. *FEBS Lett.* **336**, 389–393 (1993).
15. Verkhovsky, M. I., Morgan, J. E., Verkhovskaya, M. L. & Wikström, M. Translocation of electrical charge during a single turnover of cytochrome c oxidase. *Biochim. Biophys. Acta* **1318**, 6–10 (1997).
16. Zhen, Y. J., Schmidt, B., Kang, U. G., Antholine, W. & Ferguson-Miller, S. Mutants of the Cu_A site in cytochrome c oxidase of *Rhodobacter sphaeroides*: I. Spectral and functional properties. *Biochemistry* **41**, 2288–2297 (2002).
17. Karpefors, M., Ådelroth, P., Zhen, Y., Ferguson-Miller, S. & Brzezinski, P. Proton uptake controls electron transfer in cytochrome c oxidase. *Proc. Natl Acad. Sci. USA* **95**, 13606–13611 (1998).
18. Papa, S., Capitanio, N. & Villani, G. A cooperative model for protonmotive heme-copper oxidases. The role of heme *a* in the proton pump of cytochrome c oxidase. *FEBS Lett.* **439**, 1–8 (1998).
19. Michel, H. The mechanism of proton pumping by cytochrome c oxidase. *Proc. Natl Acad. Sci. USA* **95**, 12819–12824 (1998).
20. Rich, P. R., Jünemann, S. & Meunier, B. Protonmotive mechanism of heme-copper oxidases. *J. Bioenerg. Biomembr.* **30**, 131–138 (1998).
21. Tsukihara, T. *et al.* The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *Proc. Natl Acad. Sci. USA* **100**, 15304–15309 (2003).
22. Wikström, M., Verkhovsky, M. I. & Hummer, G. Water-gated mechanism of proton translocation by cytochrome c oxidase. *Biochim. Biophys. Acta* **1604**, 61–65 (2003).
23. Namslawer, A., Pawate, A. S., Gennis, R. B. & Brzezinski, P. Redox-coupled proton translocation in biological systems: Proton shuttling in cytochrome c oxidase. *Proc. Natl Acad. Sci. USA* **100**, 15543–15547 (2003).
24. Svensson-Ek, M. *et al.* The X-ray crystal structures of wild-type and EQ(I-286) mutant cytochrome c oxidases from *Rhodobacter sphaeroides*. *J. Mol. Biol.* **321**, 329–339 (2002).
25. Brzezinski, P. & Larsson, G. Redox-driven proton pumping by heme-copper oxidases. *Biochim. Biophys. Acta* **1605**, 1–13 (2003).
26. Siegbahn, P. E. M., Blomberg, M. R. A. & Blomberg, M. L. Theoretical study of the energetics of proton pumping and oxygen reduction in cytochrome oxidase. *J. Phys. Chem. B* **107**, 10946–10955 (2003).
27. Popovic, D. M. & Stuchebrukhov, A. A. Proton pumping mechanism and catalytic cycle of cytochrome c oxidase: Coulomb pump model with kinetic gating. *FEBS Lett.* **566**, 126–130 (2004).
28. Brändén, M. *et al.* On the role of the K-proton transfer pathway in cytochrome c oxidase. *Proc. Natl Acad. Sci. USA* **98**, 5013–5018 (2001).
29. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
30. Heller, H., Schaefer, M. & Schulten, K. Molecular-dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-crystal phases. *J. Phys. Chem.* **97**, 8343–8360 (1993).

Acknowledgements We would like to thank S. Ferguson-Miller for providing the M263L mutant strain of *R. sphaeroides* and I. Holmgren for invaluable help with preparation of Cyt cO and the SUVs.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.B. (peterb@dbb.su.se).

ERRATUM

doi:10.1038/nature04144

In situ multi-satellite detection of coherent vortices as a manifestation of Alfvénic turbulence

David Sundkvist, Vladimir Krasnoselskikh, Padma K. Shukla, Andris Vaivads, Mats André, Stephan Buchert & Henri Rème

Nature 436, 825–828 (2005)

In the print and PDF versions of this Letter, the colour scale bars in Fig. 2a, c and d should have been respectively labelled as follows: 'Proton number flux ($\text{cm}^{-2}\text{s}^{-1}\text{sr}^{-1}\text{keV}^{-1}$)'; ' B ($\text{nT}^2\text{Hz}^{-1}$)'; and 'DOP'. This labelling is correctly shown in the online HTML full text version.

ERRATUM

doi:10.1038/nature04145

Evasion of intracellular host defence by hepatitis C virus

Michael Gale Jr & Eileen M. Foy

Nature 436, 939–945 (2005)

In Table 1 of this Review Article, some reference citations are incorrect. Those in the fourth row from the top (IRF-1 regulation) should be 89,90 (and not 93,94); in the second-to-last row from top (regulation of RIG-I signalling), reference 94 should be included (to read 21,94).

CORRIGENDUM

doi:10.1038/nature04146

Action potential refractory period in ureter smooth muscle is set by Ca sparks and BK channels

T. Burdyga & Susan Wray

Nature 436, 559–562 (2005)

The panels of Fig. 3 in this Letter are incorrectly cited in the text. In the first full paragraph on page 560, these should be: line 7, 'Fig. 3b, top'; line 11, 'Fig. 3b'; line 12, 'Fig. 3c, top'; and line 13, 'Fig. 3c, bottom'. Other citations of Fig. 3 are correct. In addition, the vertical voltage scale on the bottom trace in Fig. 3d should read from -70 to 10 mV.

-  FOCUS
-  SPOTLIGHT
-  RECRUITMENT
-  ANNOUNCEMENTS
-  EVENTS

naturejobs

Breathing life into industry

One of the biggest complaints that both students and businesses have about education in the life sciences is that it does little to prepare young scientists for jobs in industry. A new programme based in Britain is now working to change that.

Developed by the South Yorkshire Bioscience Enterprise Network and funded by the European Commission and regional development agencies, the bio2work programme is a course in which graduate and postgraduate students create business and technical plans during their studies. It aims to take students with good scientific backgrounds and help them gain the commercial experience they need to hit the ground running in industry.

The eight-week summer workshop at the heart of bio2work acts as a bridge between academic and industrial science. It features speakers from local companies as well as site visits. It shows how basic research practices can synch with industrial lab standards, and teaches the students intellectual-property law and finance. It also helps the students to transfer soft skills, such as communication and information management, from

the academic bench into an industrial setting.

Last week, four students from the bio2work's first intake at Sheffield Hallam University showed that they had taken a step towards gaining the necessary background. The four — Anna Parker, Katie Lidster, Ruth Austin and Rhianna Moss — won the network's inaugural student business-plan competition.

But the award that all the graduates of bio2work await most eagerly is a job in the life-sciences industry. The programme provides a 50% wage subsidy to partner companies who hire its graduates. The real success of the course will come when local businesses hire this next generation of life-science entrepreneurs — and when such programmes become the norm, not the exception, in life-science education.



Paul Smaglik, Naturejobs editor

CONTACTS

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York
 345 Park Avenue South, 10th Floor,
 New York, NY 10010-1707
 Tel: +1 800 989 7718
 Fax: +1 800 989 7103
 e-mail: naturejobs@natureny.com

US Sales Manager/Corporations:
 Peter Bless
 Classified Sales Representatives
 Tel: +1 800 989 7718

**New York/Pennsylvania/
 Latin America:** Kelly Roman
**Midwest USA/Maryland/
 NIH:** Wade Tucker
East USA/Canada:
 Janine Taormina

San Francisco Office
Classified Sales Representative:
 Michaela Bjorkman
 225 Bush Street, Suite 1453
 San Francisco, CA 94104
 Tel: +1 415 781 3803
 Fax: +1 415 781 3805
 e-mail: m.bjorkman@naturesf.com

European Head Office, London
 The Macmillan Building,
 4 Crinan Street,
 London N1 9XW, UK
 Tel: +44 (0) 20 7843 4961
 Fax: +44 (0) 20 7843 4996
 e-mail: naturejobs@nature.com

Naturejobs Sales Director: Nevin Bayoumi (4978)
European Sales Manager: Andy Douglas (4975)

Advertising Production Manager: Billie Franklin
 To send materials use London address above.
 Tel: +44 (0) 20 7843 4814
 Fax: +44 (0) 20 7843 4814
 e-mail: naturejobs@nature.com

Naturejobs web development: Tom Hancock
Naturejobs online production: Niamh Shields

European Satellite Office
 Patrick Phelan
 e-mail: p.phelan@nature.com

Japan Head Office, Tokyo
 Chiyoda Building,
 2-37 Ichigayatamachi,
 Shinjuku-ku,
 Tokyo 162-0843
 Tel: +81 3 3267 8751
 Fax: +81 3 3267 8746
Asia-Pacific Sales Director: Rinoko Asami
 e-mail: rasami@naturejpn.com

MOVERS

Brian Foster, European regional director, International Linear Collider's Global Design Effort, Oxford, UK



2004–present: Head of Sub-department of Particle Physics, University of Oxford, UK

2003–present: Professor of experimental physics, University of Oxford, UK

2002–05: Chairman, European Committee for Future Accelerators

1984–2003: Lecturer to professor, Department of Physics, University of Bristol, UK

Believing in things of a size most people can't imagine is part of a physicist's everyday life. In his work towards the proposed International Linear Collider (ILC), Brian Foster has gone from subatomic quarks and leptons to the largest linear collider ever built.

For Foster, this is the biggest step in a career that has seen him building new equipment to take discovery further and influencing the direction of physics by advising the UK government. The ILC will be about 40 kilometres long, able to smash electrons together at 500 billion electronvolts in its search for new dimensions of space and forces of nature, and is likely to cost several billion dollars.

Foster's new role extends beyond building the machinery to constructing support for it. Some people, especially outside the particle-physics community, see the project as too big to build and too expensive to pay for. "I expect to spend a lot of time on outreach, both to fellow scientists and to the general public," he says.

Although the focus of his research has remained steady for more than 25 years, Foster has been open to change when necessary. After gaining his DPhil from Oxford, he spent nearly 20 years in the physics department at the University of Bristol, where he is now professor emeritus. Then, in his late 40s, he became chairman of the European Committee for Future Accelerators and soon afterwards took "the obvious next step" back to Oxford.

"Never get depressed when you think that your career isn't developing or advancing," he advises. "This happened to me several times. What is actually happening is that the pressure is gradually building up, more and more people are noticing you, there will be a sudden 'dam-break' and your career will move to another level."

He is always learning from colleagues: veteran physicist George Kalmus, for example, a colleague on many Particle Physics and Astronomy Research Council committees, taught him "how to keep calm and extract the best from what seemed like awful dilemmas".

He is learning from interests outside science, too — touring the globe with his violin teacher Jack Liebeck, giving a World Year of Physics lecture that celebrates Einstein (another passionate violinist) through a mixture of music and science. "Playing the violin seriously again has shown me that you are never too old to get better at something, if you want it badly enough, are lucky enough to find a superb teacher and can put in the work," he says. ■ **Janet Wright**

SCIENTISTS & SOCIETIES

Networking in the Framework

A group of PhDs and postdocs from eight European countries, all funded by the European Union (EU) as part of its Framework Programme on research, met recently in Cambridge, UK, with no supervisors, group leaders or principal investigators, but loads of enthusiasm.

The Framework Programme is the EU's main instrument for funding research. The sixth programme has a budget of nearly €18 billion (US\$22 billion). The fifth programme funded no fewer than 500 projects, each representing a Research Training Network (RTN) — a consortium of teams in different countries that propose a common project to provide training and transfer of knowledge.

Our RTN involves laboratories in Germany, the Netherlands, Italy, France, Spain, Switzerland, Denmark and Britain. Using a multidisciplinary approach, it aims to improve understanding of how genome stability is controlled by checkpoints for DNA damage and DNA repair mechanisms. It will also give young researchers intensive, internationally competitive training in a dynamic field, with emphasis on the use of integrated genetic and biochemical approaches.

The transfer of knowledge and interaction between participants is one of the key points of the network. Researchers can meet regularly without having to cross the Atlantic. Another benefit is that young scientists can

develop collaborations outside their own institutes early in their careers.

After a successful conference in Amsterdam in February, five PhDs and five postdocs involved in the project decided to meet again. The goal of this meeting, held in July in the beautiful surroundings of Magdalene College, was to provide a platform for setting up collaborations or reinforcing ongoing ones. The small numbers allowed for longer presentations and discussions, and the absence of principal investigators let the young researchers express their thoughts and ideas freely. Two important issues scientists encounter early in their careers were addressed by invited speakers: Patrick Bateson, an ethologist at Cambridge, discussed the ethics of using animals in research, and Cambridge pathologist Anne Cooke talked about writing successful grant applications.

In November 2005, this RTN will culminate in a final meeting in Milan, Italy. By that time, and thanks to the meeting in Cambridge, all young researchers involved will not only have benefited from the project and its complementary teams, but also from contacts that could prove vital to their careers.

■ **Alessandro Sartori is a postdoc at the Gurdon Institute of Cancer and Developmental Biology, University of Cambridge, UK.**

GRADUATE JOURNAL

Dualities

The first year of my PhD studies is over, and it has taught me as much about myself and others as it has about structural biology and bioinformatics. Most significantly, I have learned that most of the decisions I make have both upsides and downsides, both of which you have to come to terms with. My first year has been both a blessing and a curse. On one hand, it has been a constant struggle for money. But on the other, I feel that I am living my life the way I want.

My research experience has been equally mixed. There were moments when I felt that time was standing still and I was not progressing at all. But this just came down to the time I needed to learn new techniques to get me to the next research phase. Once there, I experienced moments of failure, followed by feelings of satisfaction, when the work moved along. I've also been very busy, commuting between Lodz and Warsaw. But I've still managed to carve out time for other non-scientific activities, such as photography and travel — which I've managed to intermingle with attending scientific conferences.

Perhaps the biggest thing I've learned in the past year is to live with both sides of a decision — and to find ways to balance the positives and negatives. ■ **Karolina Tkaczuk is a graduate student at the Technical University of Lodz, Poland.**

Not at home to visitors

They're big, but they're not clever.

Janet Wright

The guard stepped away from the small screen and said: "They're going. Come and look." The security manager hurried back, rubbing his eyes. If they really were going, he'd be glad of a good night's sleep. There was no real danger, of course — he'd kept assuring everyone of that. He'd almost convinced himself.

Intruders hadn't posed a serious threat in many years. But these were bigger than any previously seen or spoken of, and there was something different about them. He peered into the screen.

"That's a welcome sight," he said. A group of aliens was filing onto the strange craft that had caused so much interest when it arrived, followed by concern when its occupants began conducting a very thorough search. The people had learnt long ago that the best way of getting on with strangers was to avoid them. As their lifestyle made that easy, they rarely ran into trouble.

Aliens had come poking around before, but this group was using instruments the people hadn't previously seen. They were more technologically advanced than the earlier arrivals, that was for sure. There had been times when they'd almost stumbled on the ventilation system, and their heavy movements had shaken the rooms below.

They'd even found a toy that a child had left outside when she'd been playing in the moonlight the night before. She had wept over her lost birthday present, while her parents said: "It's your fault — you should have taken more care of it. You're not getting another one." Then, of course, security had sent a grab team and snatched it back for her.

The manager grinned at his guard. She was tousling the fur of a monkey that had brought another little trophy into the security room.

"You're brilliant with those monkeys," he said. The guard smiled modestly as she handed over the latest prize.

"They're intelligent and they like sweets. It's not hard to train them," she demurred. One of the monkeys had been caught but, true to its training, had gone limp until it saw its chance and leapt to freedom. "It was almost a shame seeing those aliens wondering who kept taking their equipment!"

"Yes, that was funny. The stuff itself was a bit disappointing, though, very low-tech.

I thought they might have had a few devices worth looking at. Still, it's something new for the kids to play with."

The festival director hurried into the security room. "Is it true that they're going?" she panted.

The guard ushered her to the screen. "There you are. Can't make any promises, of course — we'll need to check that they're all gone and that they haven't left any nasty surprises."

to sleep in, and just used those fabric pod things to shelter from the sky. No music or anything. And plodding around all over the place in broad daylight."

They all shook their heads and went back to work.

Lee stared at the retreating coastline. The rest of the team had never stopped mocking him since he'd screamed with joy when his trap had caught one of the creatures



JACEY

"Judging from the stuff we've seen so far, anything they've left should be easy to detect," said the security manager. "We've found a few micro-cameras already. It's not highly advanced stuff, but we'll get the monkeys to dismantle it all in case it's being monitored."

"So with luck, we should be clear by full moon then?" the festival director asked.

The others laughed. "That's the only thing that's bothered you, isn't it? That we'd have to delay the festival or hold it indoors!"

"Well, people put a lot of effort into their music, and there's no way around it, the best below-ground acoustics in the world aren't as good as outdoors on a clear moonlit night. Anyway, it's traditional. It's a pity those aliens didn't bring anything interesting," she added.

"You mean a musical instrument?" the guard teased her. "They had some slightly clever gadgets, but nothing you could describe as civilized."

"That was quite odd, wasn't it?" the security manager mused. "The aliens seemed even less advanced than their technology. They didn't even make a token little hollow

stealing their equipment. OK, not since the scream (they'd been just as excited, though they wouldn't admit it now), but since the moment their lights had revealed just a frightened little monkey.

As he'd never shown them the extraordinary astronomical device he had found, Lee couldn't admit that the monkeys had taken that — along with his illegal gun.

Two colleagues came up beside him and started waving. "Bye bye, little hobbits," they crooned. "Sorry we called when you weren't home. Say hello to Gandalf for us."

Lee turned his back on them, on the island of Flores and on the futile search for evidence of the tiny people said to have once lived there.

A few miles away, the security manager took another look at the gun the monkey had brought in. "How does this work?" he pondered. Tapping the barrel experimentally, he began composing a tune. ■

Janet Wright is a freelance journalist and author of books about health, exercise and nutrition. Sometimes her mind wanders.